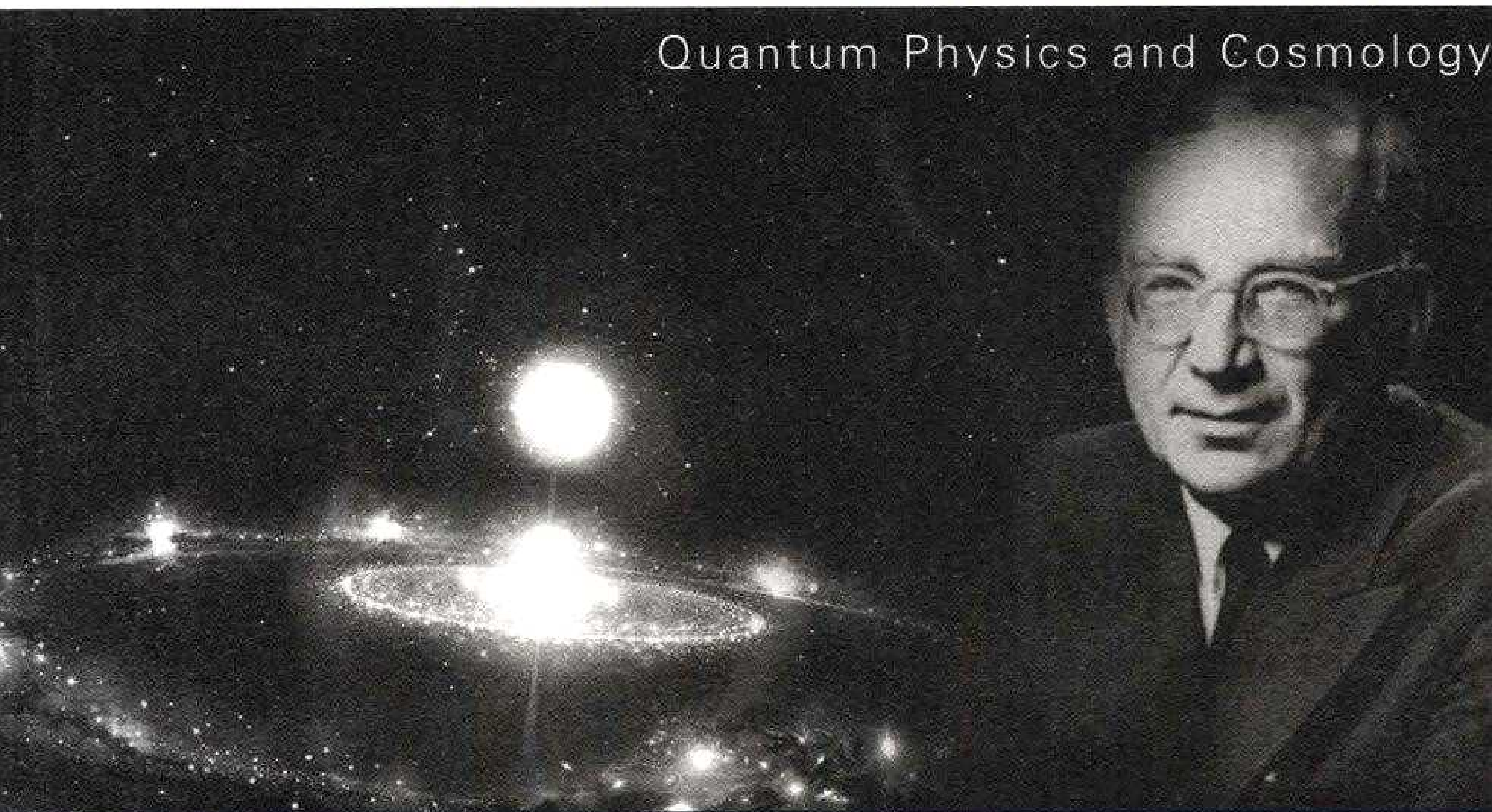


开放人文

IN SEARCH OF THE BIG BANG

Quantum Physics and Cosmology



[英] 约翰·格里宾 著　卢炬甫 译

John Gribbin

大爆炸探秘

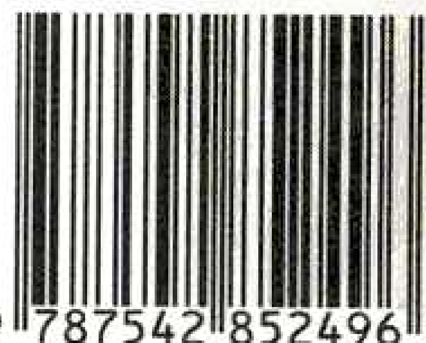
量子物理与宇宙学

上海世纪出版集团



上架建议：科普读物

ISBN 978-7-5428-5249-6



定价：51.00元

易文网：www.ewen.cc

大爆炸探秘
——量子物理与宇宙学

[英]约翰·格里宾　著　卢炬甫　译

世纪出版集团　上海科技教育出版社

图书在版编目(CIP)数据

大爆炸探秘：量子物理与宇宙学/（英）格里宾（Gribbin，J.）著；卢炬甫译.—上海：上海科技教育出版社，2011.8

（世纪人文系列丛书．开放人文）

ISBN 978-7-5428-5249-6

Ⅰ.①大… Ⅱ.①格… ②卢… Ⅲ.①量子力学-研究②“大爆炸”宇宙学-研究 Ⅳ.①O413.1②P159.3

中国版本图书馆CIP数据核字(2011)第149676号

责任编辑　卞毓麟　匡志强　殷晓岚

装帧设计　陆智昌　朱赢椿

大爆炸探秘——量子物理与宇宙学

[英]约翰·格里宾　著

卢炬甫　译

出　　版　世纪出版集团　上海科技教育出版社
　　　　　（200235　上海冠生园路393号　www.ewen.cc）

发　　行　上海世纪出版集团发行中心

印　　刷　上海商务联西印刷有限公司

开　　本　635×965 mm　1/16

印　　张　23.5

插　　页　4

字　　数　315 000

版　　次　2011年8月第1版

印　　次　2011年8月第1次印刷

ISBN 978-7-5428-5249-6/N·815

图　　字　09-2011-377号

定　　价　51.00元

出版说明

自中西文明发生碰撞以来，百余年的中国现代文化建设即无可避免地担负起双重使命。梳理和探究西方文明的根源及脉络，已成为我们理解并提升自身要义的借镜，整理和传承中国文明的传统，更是我们实现并弘扬自身价值的根本。此二者的交汇，乃是塑造现代中国之精神品格的必由进路。世纪出版集团倾力编辑世纪人文系列丛书之宗旨亦在于此。

世纪人文系列丛书包涵“世纪文库”、“世纪前沿”、“袖珍经典”、“大学经典”及“开放人文”五个界面，各成系列，相得益彰。

“厘清西方思想脉络，更新中国学术传统”，为“世纪文库”之编辑指针。文库分为中西两大书系。中学书系由清末民初开始，全面整理中国近现代以来的学术著作，以期为今人反思现代中国的社会和精神处境铺建思考的进阶；西学书系旨在从西方文明的整体进程出发，系统译介自古希腊罗马以降的经典文献，借此展现西方思想传统的生发流变过程，从而为我们返回现代中国之核心问题奠定坚实的文本基础。与之呼应，“世纪前沿”着重关注二战以来全球范围内学术思想的重要论题与最新进展，展示各学科领域的新近成果和当代文化思潮演化的各种向度。“袖珍经典”则以相对简约的形式，收录名家大师们在体裁和风格上独具特色的经典作品，阐幽发微，意趣兼得。

遵循现代人文教育和公民教育的理念，秉承“通达民情，化育人心”的中国传统教育精神，“大学经典”依据中西文明传统的知识谱系及其价值内涵，将人类历史上具有人文内涵的经典作品编辑成为大学

教育的基础读本，应时代所需，顺时势所趋，为塑造现代中国人的人文素养、公民意识和国家精神倾力尽心。“开放人文”旨在提供全景式的人文阅读平台，从文学、历史、艺术、科学等多个面向调动读者的阅读愉悦，寓学于乐，寓乐于心，为广大读者陶冶心性，培植情操。

“大学之道，在明明德，在新民，在止于至善”（《大学》）。温古知今，止于至善，是人类得以理解生命价值的人文情怀，亦是文明得以传承和发展的精神契机。欲实现中华民族的伟大复兴，必先培育中华民族的文化精神；由此，我们深知现代中国出版人的职责所在，以我之不懈努力，做一代又一代中国人的文化脊梁。

上海世纪出版集团

世纪人文系列丛书编辑委员会

2005年1月

应当相信的是观测而不是理论，理论之可信也只是在被观测所证实的限度内。

——亚里士多德

我看到的越多，我看到有待去看的也越多。

——约翰·塞巴斯蒂安

内 容 提 要

茫茫宇宙始终是人类赞叹和歌咏的对象。可是，宇宙从何而来？它是否永远如此？千百年来，人们总是把这些问题留给宗教和哲学，不是归功于上帝的“第一推动”，就是将其当作形而上学者苦思冥想的话题。然而今天，科学家们正在努力构建一幅宇宙诞生与演化的生动图像：宇宙起源于100多亿年前的一个“原始火球”，或曰“大爆炸”，在经过极其猛烈的暴胀之后，不断膨胀和演化，直到形成现在这个丰富多彩的大千世界。

从托勒玫、哥白尼、赫歇尔到哈勃，无数天文学家的不懈努力使我们的目光从地球扩展到太阳系、银河系、河外星系，乃至我们生存于其中的整个宇宙。本书不仅将带您领略这一激动人心的科学历程，其笔触更深入到现代物理学的最前沿，生动地描述20世纪物理学的两大奇葩——由爱因斯坦创立的广义相对论和由普朗克、玻尔、海森伯等人奠基的量子物理学，如何在20世纪后半叶汇聚于宇宙学研究之中，并最终为我们初步揭开了宇宙的创世之谜。

本书追踪了大爆炸理论的来龙去脉，不但向您展示了现代宇宙学的巨大成就，更让您身临其境地接触那些伟大的科学先驱者们，正是站在他们肩上，我们才有可能看得如此真切、如此深远。

作 者 简 介

约翰·格里宾，英国著名科学读物专业作家，萨塞克斯大学天文学访问学者。他毕业于剑桥大学，获天体物理学博士学位。曾先后在《自然》杂志和《新科学家》周刊任职。1974年他以其关于气候变迁的作品获得了英国最佳科学著作奖。

约翰·格里宾著有50多部科普和科幻作品，其中的科学三部曲《薛定谔之猫探秘》、《双螺旋探秘》和《大爆炸探秘》尤为脍炙人口。此外，他还与妻子合著了一系列著名科学家的传记，而反映“科学顽童”费恩曼科学生涯的《迷人的科学风采——费恩曼传》(中译本于1999年由上海科技教育出版社出版)更是广受好评。

致　谢

本书的根源要回溯到很久以前，回到20世纪50年代初我开始对科学产生兴趣的时候。我不太想得起来是哪位作者最早带我进入宇宙的奥秘和奇迹，但我知道此人如果不是阿西莫夫(Isaac Asimov)，就一定是伽莫夫(George Gamow)，因为我很早就开始读他们的书，我简直不能想象会没有他们。而且不仅是科学吸引了我，宇宙的起源之谜更是从一开始就令我沉醉。由于伽莫夫和他杜撰的“汤普金斯先生”，我开始知道了宇宙起源的大爆炸模型。虽然后来我接触了稳恒态假说，但大爆炸的思想——存在一个产生宇宙的确定时刻的思想——始终占据着我的心灵。我从未想过将来会不去研究这些深奥的难题，或者写些关于它们的书。实际上，直到1966年，我都一直没有认识到，要做一名天文学家，更不用说一名宇宙学家，并不是件谁都能做的工作，更别说是我了。此后，正当我在萨塞克斯大学参加最后的大学考试时，我发现麦克雷(Bill McCrea，现在是麦克雷爵士)将在校园里建一个天文学研究中心。

这个发现改变了我的生活。首先使我立即将读粒子物理研究生的计划改变为在麦克雷的小组读天文学硕士。后来我到了剑桥，成为另一个新的天文学小组——霍伊尔(Fred Hoyle，现在是霍伊尔爵士)那时的理论天文学研究所——非常低级的最初成员。由于某些原因，我并没有非常投入，却转而把对极致密恒星(白矮星、中子星、脉冲星和X射线源)的研究作为我的学位论文，而基本上没有做过什么宇宙学方面真正的工作。但在剑桥我见到了霍伊尔本人，纳里卡(Jayant Narlikar)，里斯(Martin Rees)，伯比奇夫妇(Geoffrey & Margaret Burbidge)，霍金(Stephen Hawking)，福勒(William Fowler)和其他著名天文学家，他们都专注于实实在在的宇宙学重要问题的研究。从他们那儿我知道了这个层次的研究是什么样的，还知道我本人根本没有指望取得什么值得一提的成就。于是我成了一名作家，不单是报道宇宙学和天文学方面的最新进展，而是关于整个科学，不断保持与最新进展的接触，即使我本人并未置身其中。

当天文学在20世纪80年代产生巨大飞跃的时候，它是来自同粒子物理的联姻，而在1966年我曾轻率地放弃了这方面的工作。在最初竭力应付那些看起来出现得如此之快以至于我都来不及写下它们的新进展之后，我很幸运地抓住了一个机会，以旁观者的身份参加一个1983年11月在日内瓦召开的由欧洲南天天文台(ESO)和欧洲核子研究中心(CERN)共同组织的会议。在那儿，来自粒子物理学和宇宙学方面的参加者们讨论了其中的联系。正是那次会议，以及对自己能了解那儿所发生的绝大多数事情的自信，使我确信自己能写这本书。由于这次会议，我能够理清我的思路，并在同CERN的纳洛普罗斯(Dimitri Nanopoulos)及暴胀假说的两位建立者麻省理工学院的古思(Alan Guth)和莫斯科的林德(Andrei Linde)通信后，加深自己对暴胀

这一大爆炸宇宙学现代版本的关键性新思想的理解。

我于1985年夏写下这些话的时候，看起来科学已经获得了一个(至少是在概貌上)完整的理解，知道我们所知的宇宙如何产生和它如何通过大爆炸从一粒小种子成长为我们所见的如此广漠。剑桥大学的里斯已经清楚地说明了这项新工作的重要性。他在1983年11月的那次日内瓦会议上评述说，当被问到大爆炸是不是对我们生活于其中的宇宙的一个好的模拟时，他过去总是说："这是我们已经得到的最佳理论。"这确实是十分谨慎的认可。但现在，他在日内瓦说，如果现在被问及同样的问题，他会回答说："大爆炸模型更有可能被证明是对的，而不是错的。"里斯是宇宙学家中最为谨慎者之一，他从不轻易下结论，他的话是对大爆炸的强烈支持，也给了我足够的理由来写这本书。

我能了解这些新思想背后的物理，这要归功于我学生时代和在萨塞克斯及剑桥时的老师们的技巧。能够生活在这些奥秘被揭开的时代，能够了解它们如何被揭开，是我所能想象的一种最大的幸运。也许会有新的奥秘来扰乱目前的这幅图像，而完全了解创世时刻也可能被证明只是一个梦幻。但今天的图像已足够完备，我希望通过本书与您分享对其完备性的惊叹，和在发现宇宙膨胀因而必须存在一个创世时刻后的60年对一个成功的创世理论的探索。

如果我最终成功地吸引了您的注意力，这主要是由于故事本身是如此迷人，只有一个拙劣的讲述者才会令它毫无生气。我要感谢阿西莫夫和伽莫夫，他们向我讲述了这个传奇的早期版本；感谢麦克雷，他出现于萨塞克斯大学的校园里，向我展现了宇宙学家也是活生生的人，而我也能和他们一同工作；感谢霍伊尔，他建立了一个研究所，使我有可能接触一流的宇宙学家们；感谢CERN邀请我参加第一次

ESO/CERN讨论会，同时也感谢《新科学家》派我去报道那次会议。在本书写作过程中，我得到了来自古思、林德、纳洛普罗斯、剑桥的里斯以及孟买塔塔研究所的纳里卡的直接帮助。麦克雷在其忙碌的生活中抽空阅读本书草稿的头两部分并纠正了我的一些历史误解，牛津大学的克洛斯(Frank Close)对有关粒子物理学的章节也作了同样的努力，而里斯指出了我对宇宙学不够了解的地方。但无疑仍会有遗留的错误，这完全是我的责任。如果您发现了，请让我知道，我会尽力在以后的版本中纠正。但我希望这些错误很少、很小，不会妨碍您欣赏这个探索宇宙最终真相和宇宙自身起源的故事。

约翰·格里宾

1985年6月

目录

序幕　玄奥的宇宙

人们总想知道关于创世的奥秘。我们今天所称的宇宙是怎么来的？为什么它能存在？我们又怎么能在这里作此深思？这些问题在数千年中一直属于哲学和神学的领域。即使在四五个世纪前现代科学奠基之际，某些问题超出了科学的范围，这一信念仍被普遍接受。例如，牛顿(Isaac Newton)力图解释存在于宇宙之中的物体的行为，却从不涉及宇宙本身的由来。然而今天，科学正在进入哲学的领地。现代物理学即量子物理学，正在回答诸如存在的本质、生命的本质，乃至整个宇宙的起源这样的终极问题。

是科学变成了哲学，还是哲学变成了科学？无论你如何看待，可以肯定的是，这二者之间的界线已经变得模糊，变得远不如今天大多数科学家和哲学家自己所认为的那样实在。令人惊讶的是，今天被纳入物理学的这部分哲学，是哲学中最为幽深玄奥的，被称之为形而上学，其源可上溯到两千多年前的亚里士多德(Aristotle)的巨著。所以，尽管这本书是用现代天文学和物理学知识来讲述宇宙的起源，讲

述宇宙如何诞生于原初火球即所谓大爆炸，看来还是有必要先回到亚里士多德，并对那些已被形而上学家们思索了许多个世纪的难题来一番浏览。

亚里士多德于公元前384年生于希腊北部的马其顿。他父亲是马其顿国王亚明塔斯三世（Amyntas Ⅲ）[亚历山大大帝（Alexander the Great）的祖父]的宫廷医生。亚里士多德40岁出头时曾应国王菲利普二世（Philip Ⅱ）之聘担任少年王子亚历山大的教师。但是亚里士多德在今天的大名并非由于他在马其顿宫廷的地位。他的大量科学和哲学著作中，有两本与现代对宇宙本质的探索密切相关。一本是《物理学》，论述可见的世界。另一本是《形而上学》（字面的含义是“物理学之后”），探究潜藏的、支配可见世界的真实，即亚里士多德所称的“本原”。

按照现代的划分，这两本大作都属于哲学著作，但是亚里士多德所试图作出的区分，即我们所见的或者说可用科学仪器来测量的世界与潜在的真实之间的区分，是非常重要的，它直指现代物理学的核心。在亚里士多德的继承者中对“本原”的研究被称为形而上学，它在许多个世纪里被哲学家看作最基础的学科，因为它所探求的是对真实的整体性、综合性认识。

这似乎使人觉得难以把形而上学家看作典型的哲学家，因为后者往往更致力于对事物的精细分析，而不是对世界本质的探究。有些形而上学家甚至怀疑一棵树或一间房子在没人看它时是否还真实存在，这在我们这些凡夫俗子看来简直是太可笑了。但可笑的是我们，因为，最实际和最客观的科学——物理学，在20世纪的发现已经不可抗拒地导致这样的结论，即在亚原子粒子诸如电子和质子的层次上，物体在未被观测时的确并不“真实”存在。

这种对真实的终极本质的关注是形而上学的三大根源之一。形而上学家想要知道的，正是我们感官提供的关于真实世界的图像是否准确。我们的感官对从外部世界接受的印象产生反应，我们的大脑解释这些感官印象并指出(比方说)在花园里有一棵树。但是我们的大脑能直接知道的仅仅是感官印象，而我们对树的“知道”则是第二手的，是经过感官过滤而进入大脑的。那么，哪个更真实——是感官印象还是树？我不准备在这里从传统哲学上详述这个问题，我要强调的是，上述讨论与20世纪20年代以来物理学家关于由实验探测的亚原子粒子世界真实性的讨论惊人地相似。从来没有人看见过电子或是原子。我们推断存在着我们称为电子或原子的东西，是因为一旦我们做某些实验，就能得出与电子和原子的存在相符的结果。但是我们确切“知道”的只是对仪表读数或荧屏上闪烁光点的感官印象，而从来不是对我们自信正在研究的粒子的直接感觉。著名的物理学家马赫(Ernst Mach，他的贡献之一是用于衡量飞行器速度的物理量马赫数)于1883年在他的《力学》一书中曾这样概括：“原子是不可能被我们的感官觉察的，它们像所有物质一样只是思想的产物……只是一个便利于事实在思想上再生的数学模型。”

50年后，另一位大物理学家阿瑟·爱丁顿爵士(Sir Arthur Eddington)在他的《物理世界的本质》一书中也考虑了由科学仪器所报告的现实之本质这个问题。他指出，一张桌子对我们的感官来说表现为一样实实在在的物件，但是按照我们所钟爱的物理定律，那些探测物质结构的实验结果只能表明那桌子几乎是真空，它实际上是由许多很小的原子组成，而原子之间隔着比它们自身大得多的真空。阿尔伯特大厅中央的一粒沙，还不如那桌子里的一个原子那样孤独。那把原子束缚在一起组成桌子的看不见的电磁力倒真是“实实在在”的。

感官印象中的桌子，比起由科学研究所得知的组成桌子的原子来，是更真实还是更不真实呢？所有这些已经够令人困惑的了，然而更糟糕的还在后头。

当物理学家试图探测电子和组成原子的其他粒子的精确性质时，他们发现那“真实”粒子的概念溜掉了。为了解释电子、质子和其他粒子的行为，他们不得不在20世纪20年代发展出一种新型的物理学，即量子物理学。“物理学之后”正是量子物理学，这样说并不夸张；更奇妙的是，量子物理学又正是用来解开形而上学疑难的法宝。在这个粒子物理的新世界里，粒子和波原来只是同一事物的两个方面。光曾被认为是电磁波，现在则必须被看作还是一种叫做光子的粒子流；电子曾被看作是像小弹子球那样的粒子，现在则必须被认为也是一种波。更糟的是，当20世纪20年代的物理学家试图运用量子物理的新认识来预测电子或其他粒子在实验装置中的行为时，他们发现这是不可能做到的，除非是在统计意义上。按照量子物理学，当你不看一个电子时，你无法知道它在哪里、在干什么。如果你测量一个原子的某种性质并得到结果A，那么你所能做的一切只是计算下一时刻你作同样测量而得到结果B的概率。可是当你真的测量时，又会有一定的概率得到不是B而是C的结果。

这种状况听起来的确使人无所适从，以至于今天的大多数科学家和工程师都对它置之不理，而继续把电子当做小小的、可预测的弹子球，尽管他们用来设计激光或核反应堆的方程式是基于20世纪20年代得出的那些古怪的量子物理定律。量子物理学已经提供了形而上学三大难题之一的“答案”，那就是，如我在《薛定谔之猫探秘》一书中所讲的，按我们对“真实”这个术语的日常理解，没有什么东西是真实的。

所以，量子物理学对付的基本难题是，物体当你不看它们时是什么样的，以及即使你在看时它们又是否真是实在的。形而上学三大难题之二是关于生命的起源和本质，现在也能依据量子物理学来理解，因为这门新物理学提供了原子之间相互作用的基本定则，正是这些定则允许包括双螺旋生命分子 DNA 在内的分子存在。同关于真实本质的难题一样，量子物理学与生命奥秘的联系也得用一整本书来讲，于是我又写了一本《双螺旋探秘》。上述两个难题不在这里进一步讨论了，因为现在我要着手对付第三个也是最大的一个形而上学难题：这一切怎么会存在？我们的世界亦即宇宙是从哪里来的？

20 世纪 20 年代发现的量子物理学基本定则，同样也对这一难题作了洞察，尽管当时没有人意识到这一点。物理学家已经用了 60 年来领会那些定则的形而上学含义，直到 80 年代才首次比较令人满意地解释了宇宙是怎么演变成今天这个样子的，它又是怎么开始出现的。自 20 世纪 50 年代以来，大爆炸理论已成为“最流行”的宇宙学理论，它认为宇宙起源于一种极高密度和极高温度状态的爆发。但是，大爆炸本身又是怎么来的呢？这是现在正在回答的问题，我希望在本书里告诉你人们是如何得出答案的。20 世纪 20 年代以后，对大爆炸的探索是沿着两条分离的途径进行的，这两条途径最近才会聚成现在的答案。量子物理学家继续忙于研究原子和粒子世界的性质，却很少(如果还曾有过的话)抬眼看看天空，并想想他们的小世界是否还能告诉我们一点这个大世界的由来。与此同时，天文学家在沿着另一条途径前进，那条途径与形而上学沉思的关联并不亚于量子物理学家关于真实本质的难题。

从柏拉图(Plato)到康德(Immanuel Kant)到现在，哲学家们在苦苦思索时间和空间的本质。这个问题虽被公认为不及真实性问题或万

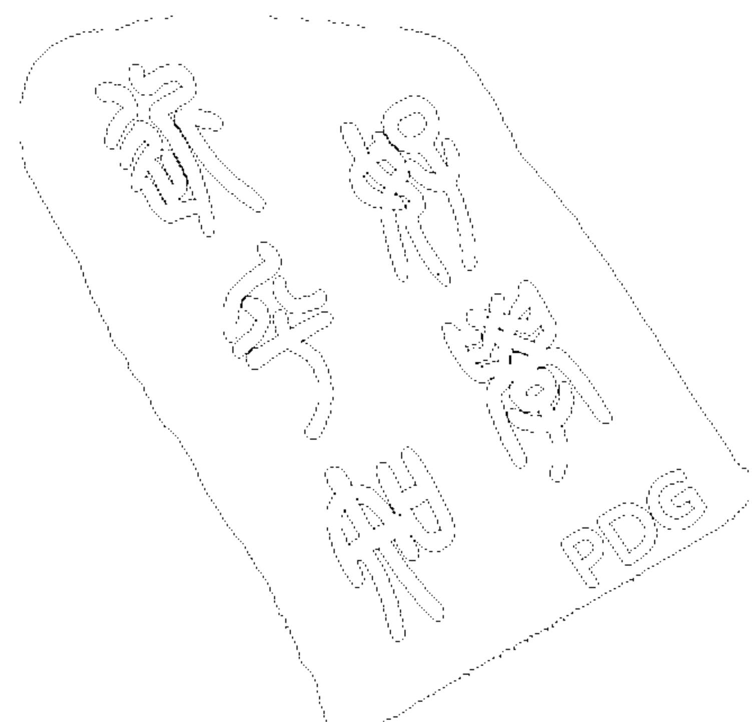
物之源问题那样难，但也击中了要害。我们都知道，必须确定空间和时间这二者，才能准确地标记一个事件。如果你安排“6点在钟楼下”见某人却没说是哪个钟楼，或者指定了钟楼却没说是什么时候见，显然都是不妥的。但是时间和空间比原子和电子更真实吗？它们会不会只是我们感觉的产物呢？试图从科学上建立认识时间和空间以及二者相互关系的框架的人，就是爱因斯坦（Albert Einstein），他也正是由此引导科学家们走上了通往大爆炸之路。

第一篇

爱因斯坦的宇宙

宇宙的最不可理解之处正在于它是可以理解的。

——爱因斯坦


PDG

第一章　星云世界

我们生活于其中的宇宙很大，并且几乎是空的。像太阳那样的明亮恒星聚集在一起组成星系，一个星系可以包含上万亿颗恒星。但只要看看那黑暗的夜空，并想想那每一个小光点本质上都是一颗类似于太阳的明亮恒星，我们就能明白恒星之间相距是多么遥远。在没有月亮的夜晚和远离城市灯光的地方，用肉眼能看到的恒星不超过 2000 颗。组成我们这个星系的绝大多数恒星，则聚集在我们称为“银河”的朦胧光带上，它们是那样遥远因而显得那样暗弱，不借助望远镜就无法区分它们。

银河系本身又只不过是宇宙汪洋中的一个小岛、茫茫太空中的一个小点。正如银河系里有着千百万颗恒星那样，宇宙空间散布着千百万个星系岛，它们相互之间的距离比银河系本身要大上许多倍。

对于研究整个宇宙的性质和演化的宇宙学家来说，星系是他认为值得考虑的最小单元。而对于一个普通人来说，银河系这个星系恐怕已经是他凭自己的感官，而且只是一种感官——视觉，所能认知的最

大限度。因为他是生活在这个像磨盘般转动的星系外部区域里一颗普通恒星主宰下的一颗小小行星上。

现在对宇宙的最好的科学认识是，它在一团火、一次发生于大约150亿年前的大爆炸中诞生。宇宙学家已经能够解释宇宙如何从一个超高温、超密集的火球，演变成现在这种广漠空间中散布着星系岛的状态。他们也至少能推测大爆炸本身是如何发生、为什么会发生，这应归功于剑桥的霍金(Stephen Hawking)等学者的最新研究结果。他们还能至少是大致地预言宇宙的最终结局。介绍这些关于整个宇宙的起源和终结、关于存在着的和我们所能认识的一切的知识，正是本书的任务。但是，假如没有银河系只是千百万个星系中的普通一员这一发现，所有这些知识就都不可能得到。对宇宙起源即对大爆炸的科学探索，开始于其他星系被确凿地证认为也像银河系那样的恒星集团之时，这种证认是在20世纪20年代做到的。

宇宙学真正是一门20世纪的科学。但像20世纪的所有其他科学一样，它的根源可回溯到古代自然哲学家和形而上学家的沉思冥想。

最初的理论

古希腊人和古罗马人认为，地球是宇宙的中心和最重要的成员。尽管希腊哲学家已经对月亮的距离有了很不错的认识，但直到17世纪望远镜出现，人们才开始领会恒星的遥远。伽利略(Galileo)是把望远镜用于天文观测的第一人。他惊讶地发现，即使借助于望远镜的放大功能，恒星看上去仍然只是光点，而不是太阳和行星那样的圆球。这只能表明恒星比太阳和行星遥远得多。他用望远镜看到了许多用肉眼不能看到的恒星，还揭示出那道被称为银河的光带是由大量密集的恒星组成的。与伽利略打开观测宇宙的新窗口同时，即在17世纪初

期，开普勒(Johannes Kepler)正在建立关于我们的小庭院即太阳系的理论体系的基础。他发现的行星绕太阳公转一周的时间与该行星到太阳的平均距离之间的关系，到 17 世纪 70 年代被用来相当精确地估算出地球与太阳的距离，这个距离的现代测量值约为 1.5 亿千米。开普勒的观测还是牛顿的引力理论的基石之一。

天文学家又用了 150 年从观测和理论两方面努力改进，到 19 世纪 30 年代末才首次测定几颗恒星的精确距离。这一测定以及进入 20 世纪后对更多恒星的距离测定，对于测量最遥远星系的距离从而确定宇宙的尺度是至关重要的。但是即使在恒星的距离被精确得知之前，17 世纪的发现已经展现了对宇宙的新视野，比起古代的那幅宇宙是一系列环绕地球的水晶球、外延只不过稍稍超出土星轨道的图像来，的确是堪称革命性的了。18 世纪有几位哲学家用一个想象的关于银河系及其在宇宙中位置的模型来解释那些新发现。那个模型竟令人惊讶地接近于现代的见解，并且在天文学家和哲学家中引发了长达两个世纪的争论。

关于宇宙的新理论——第一个近代的宇宙学理论，应归功于赖特(Thomas Wright)，英国达勒姆的一位出生于 1711 年的哲学家。与同时代的大多数思想家一样，他的学识涉及许多学科，其中之一是天文学。他的父亲是一个木匠，他对天文学的兴趣是由幼年时的教师引发的。但是，由于严重口吃，他没有能正常地受教育，有一段时间他变得很野，用他自己后来的话说，“终日游荡”。13 岁时他去跟一个钟表匠当学徒，一连干了 4 年。他把空余时间全用来自学天文学，母亲很支持他这样做，父亲却坚决反对，甚至烧掉了他的书。步入成年，赖特又连受挫折。他先是去当水手，第一次出海就遇上大风暴，再没干过第二次；接着去桑德兰当家庭数学教师，又与一位牧师的女儿卷

入了绯闻；后来又给水兵教过航海学，还上过出版商的当，想出一本历书却没搞成。到了18世纪30年代，他终于时来运转，当上了贵族的家庭教师和顾问。口吃也许仍有，但已不再是这个满怀自信的青年的障碍。他进入了人生的新阶段，讲授着自然哲学、数学、航海学等私立课程，并开始出版自己的著作。到1742年，他已颇有名气，因而被圣彼得堡的皇家学院聘请当航海学教授，年薪300英镑，但他谢绝了，因为他要求将年薪增至500英镑而未获同意。1750年，他的著作《最初的理论或关于宇宙的新假设》出版，这使他生前既享有成功、博学（主要靠自学成才）的哲学家的盛名，今天也仍被人们记起。这本书在科学史上占有重要的地位，因而在1971年被摹制重印。

18世纪*初，即伽利略的革命性发现100年后，自然哲学家已经普遍认同，恒星必定是像太阳那样自己发光的遥远天体，而不是像月亮那样反射太阳光。恒星既然不能被望远镜显示成圆盘状，它们就必定非常遥远。赖特的许多（但不是所有）同时代人认为，既然我们所知道的最明亮天体是太阳，那么关于恒星的最好猜想应是，它们也是像太阳那样明亮的天体，坐落在远处。有些人推测，恒星在空中分布的表观无规则性，只是由于我们自己身处一个系统之中，而这个系统的形状和结构只有从其外面才能看清，这就像一个森林里的人弄不清整片森林是什么样子，而只能看到在各个方向树木的无规分布一样。

如霍斯金（Michael Hoskin）在为《最初的理论》1971年版所写的前言中介绍的，赖特在其早期著作中曾认为，（我们现在所称的）宇宙是一个球体，其中满布着恒星。他还具体描绘了一个身在像地球那样环绕某一恒星运行的行星上的观察者，将如何看到附近的恒星明亮而

---

* 原文为17世纪，疑误。——译者

清晰，远处的恒星则暗弱得不借助望远镜就无法看见，更遥远的恒星则混合成一条横贯天空的朦胧光带。这番描绘似乎与银河的状态相符。但是，如赖特在后来(假如不是马上)意识到的，如果我们周围各个方向上的恒星都是均匀分布的，那么整个天空就应该都像银河光带那样亮。在1750年的《最初的理论》一书中，赖特才说中了银河表现为天空中一道光带的缘由。他写道，必须设想所有的恒星“都以同样方式运动，都在同一个平面附近，就像行星在各个同心圆周上绕太阳转动一样”。这就是说，太阳只是这样一大群恒星中的一员，它们聚成一个平盘，而不是一个球体。当我们沿着盘面朝所有恒星绕之旋转的中心看去时，看到的是大量恒星密集而成的我们称之为银河的光带。而当我们朝这个薄盘上下两侧的太空深处看去时，看到的就只是少量邻近的恒星而不是光带，因为并没有很远的恒星可见。实际上，赖特的恒星系统更像土星的有中央空隙的光环，而不是一个完整的恒星盘。坦率地说，他关于银河本质的推测，并不如他自己所认为的是对科学和哲学的最重要贡献，因为他的主要兴趣集中于今天认为本应属于宗教领域的课题。今天看来他所作的最深刻洞察，倒是他在书末的概括中几乎作为题外话说出来的。在作出对银河形态的解释之后，他继续推测在这个空间岛的外部可能有什么。他把我们今天所称的银河系称作宇宙，或是创世，并且想象所有恒星都各有自己的行星家族，都组成太阳系那样的体系。他写道：“就像可见的宇宙里满布着恒星系统和行星世界一样，无边的太空里也有着无数个与已知的宇宙类似的宇宙。”这就是说，赖特想象在银河系以外的太空还分布着许许多多像银河系那样的星系。他甚至提出，天空中那些被称作星云的、用望远镜可以看到却不能被分解成恒星的暗弱光斑，“可能是外部的宇宙，[它们]与已知的这个[银河系]类似，但是太遥远，而非望

远镜的能力所及”。

赖特的这些话，堪称是认识银河系和宇宙本质的思想火花。这火花渐渐燃成熊熊火焰。首先为之助燃的，是另一位哲学家康德和另一位天文学家威廉·赫歇尔(William Herschel)。

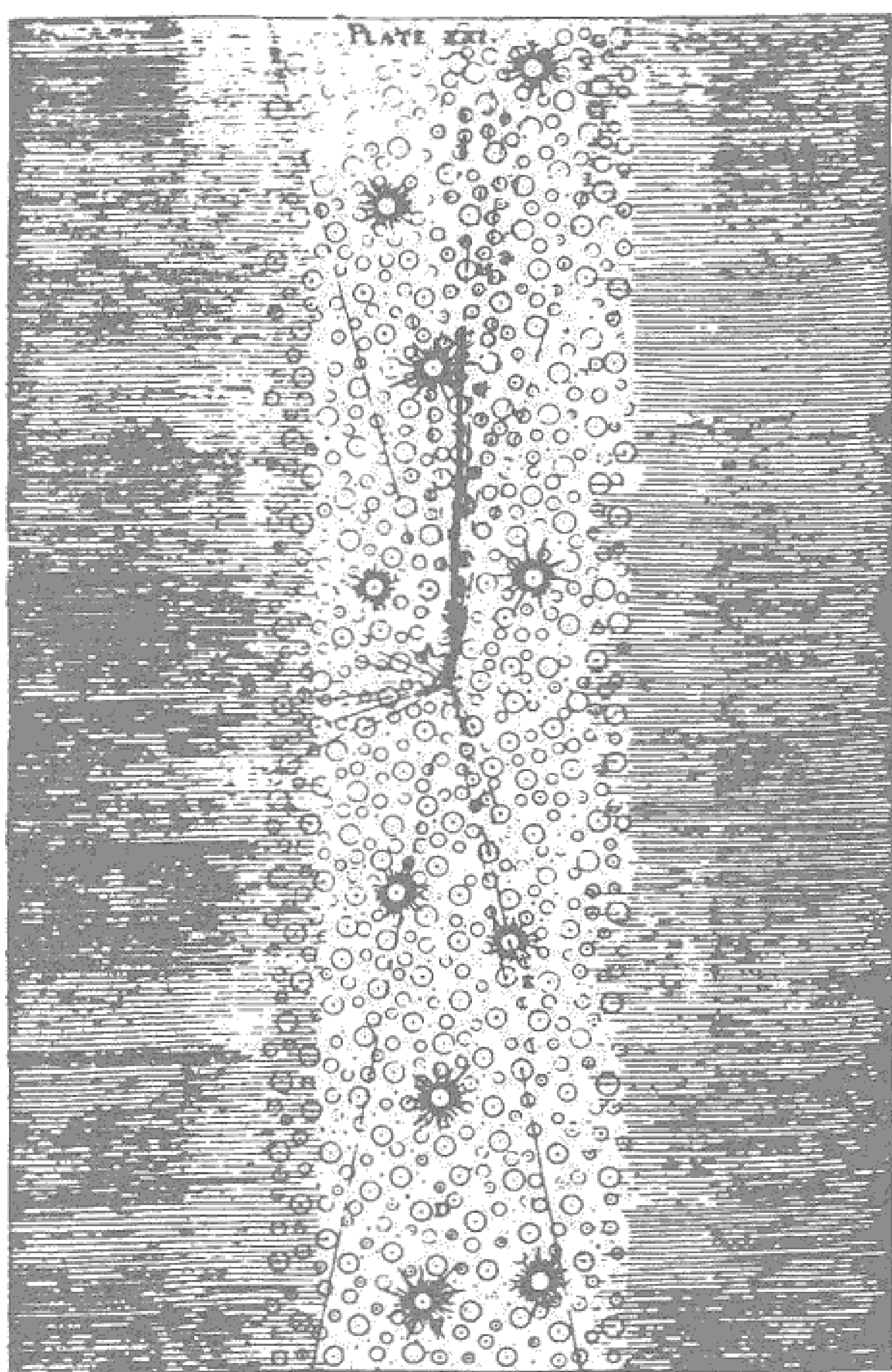


图 1.1　赖特解释银河是恒星盘的示意图，取自 1750 年出版的《最初的理论》。

星云假说

康德主要是作为一位第一流的哲学家而大名赫赫。他的教导在过去的 200 年里影响着所有的哲学思考。即使如此，《不列颠百科全书》关于他的条目里几乎不提他对天文学的兴趣仍然令人惊讶，更令人惊讶的是竟只字不提他于 1755 年出版的早期著作《自然通史和天

体论》。康德显然深受5年前面世的《最初的理论》的影响，但他在自己的书中表述的关于宇宙本质的推测要比赖特的清楚得多，而且先知似地与建立在20世纪观测基础上的现代宇宙绘景非常相似。康德从未做过任何实验或观测，他只是思索别人所作出的发现的含义，因而在许多年里被人嘲讽为“空想学者”。像他这样单凭智慧而非实际技能的“空想学者”实在不可多得，爱因斯坦便是另外一位。

康德于1724年出生于当时东普鲁士的柯尼斯堡(即以后苏联的加里宁格勒)*。他的祖父是来自苏格兰的移民，原姓坎特(Cant)，定居后把姓氏日耳曼化了。他父亲是制马鞍的工匠，生了11个孩子，家庭的贫困可想而知。伊曼纽尔·康德本来排行老四，但在幸免早夭的孩子中是最年长的。他的父母是虔诚的路德教徒，正是牧师的影响使伊曼纽尔得以从小受到正规教育。1740年他成为柯尼斯堡大学攻读神学的学生，但是真正吸引他用功学习的课程是数学和物理。他本可继续自己的学术生涯，但当父亲于1746年去世时，他不得不离开大学去当一名私人教师。这倒并不是一个不适宜的工作，因为他在此后9年中的3位雇主都在社会上颇有影响，他们使他有了一种新的生活方式和一个新的社交圈子。这期间康德作了他一生中最远的一次旅行，到了相距60英里(约100千米)的安斯多夫城。

1755年，他终于成为柯尼斯堡大学的一名教师，但是并无固定工资，报酬直接来自学生的学费。尽管别处的大学愿为他提供更好的职位，他却终生留在家乡。他在1770年成为逻辑学和形而上学教授，当了27年，后于1804年辞世。那本将使他永远留在宇宙学家记忆里的著作，是在他还只是一名私人教师时完成的，并在他首次得到大学

---

* 今属俄罗斯。——译者

职位的那一年出版。

赖特的《最初的理论》从未广泛流传，而是很快就变得罕见了。据我们所知，康德本人从未见过那本书，他对赖特思想的了解是通过汉堡的一家杂志于1751年刊登的一篇很长的书评。那篇引用了赖特一些原文的书评写得不仅很准确，而且甚至比赖特的原书还要清楚，因为它突出了赖特的核心思想，其中就有银河是“都以同样方式运动，都在同一个平面附近”的许多恒星的集合的思想。康德在自己的书中发挥了这一思想，并且对星云的本质作了比赖特有力得多的阐述，那些被称为星云的暗弱光斑能用望远镜看到，而又显然不是单个恒星。20世纪宇宙学的先驱者之一哈勃(Edwin Hubble)在他的《星云世界》一书里对康德《天体论》中的关键一段作了很好的翻译：

> 现在来讲我的体系的另一部分。这部分在我看来是最有魅力的，因为它提出了关于创世方案的崇高思想。把我们引导到这一思想的程序是非常简单和自然的。让我们想象一个聚集在同一平面上的恒星系统，它很像银河，但离我们很远，以至于用望远镜也不能分辨出组成它的恒星……这样一个恒星系统对遥远的观测者将只表现为一个光斑，既很暗弱，又只有很小的张角。如果该系统的平面与观测者视线相垂直，它看上去就是圆形的；如果相倾斜，就是椭圆形的。它的暗弱光芒、形状和可见直径，显然使它能与单个恒星相区分。
>
> 我们已无须再去费力寻找，上述的奇怪现象已经被不少天文学家看到了，并且使他们感到困惑。[1]

哈勃对康德的论证十分赞赏，因为它依据的是今天所称的一致性

原理，或者稍微拗口一点，地球平庸原理。这个原理说的是，我们生活在宇宙中一个典型的、普通的部分，宇宙的其他任何部分也同样典型和普通，看上去同我们这部分别无二致。康德的意思正是，我们生活在一个普通的星系里，而那些被天文学家看到的暗弱星云只不过是与我们这一个星系同样普通的星系，它们并无新奇之处。如果那些星系的某一个里有一颗恒星周围的一个行星上也存在智慧生物，那么银河系也就成了他们只有用自己的望远镜才能看到的暗弱光斑。首次清楚地表述这一思想的康德似乎犹嫌不足，他还首创了“宇宙岛”这个词来描绘星系散布于广漠太空的状态。

康德和赖特(后者的著作在200年里只是通过被前者用作参考文献才为世人所知)获得了荣誉，但这是由于机缘巧合。兰贝特(Johann Lambert)，一位博学的瑞士日耳曼人(其贡献之一是首先严格证明了圆周率 π 是一个无理数)，也独立地提出了一个类似的宇宙方案并于1761年发表。根据他写给康德的一封信(我们没有理由怀疑他信中的话)，他的思索开始于1749年，并且是在自己的文章完成之后才得知赖特和康德的著作。但是，自18世纪60年代往后，宇宙岛的思想在天文界就变得广为人知了，尽管尚未被普遍接受。凭借着对天空已有的很有限的观测，空想学者们已经让自己的思想驰骋很远，现在应该是观测家再次占据天文学舞台中心的时候了。但直到20世纪，观测才足以最后证实星云本身也是星系的假说。此后就一发而不可收，如洪水般涌来的观测资料使理论家们应接不暇，再不像他们的前辈那样只凭一叶而知秋。

观测宇宙

18世纪的观测天文学家中几乎没有人对星云的本质感兴趣，他

们更关注的是彗星。发现彗星，除了其科学意义外，还能使发现者一举成名，因为他的名字传统上会被用来给“新”彗星命名，当时如此，现在依然。哈雷(Edmond Halley)这个名字永远与彗星相连，他断定在1456年、1531年、1607年和1682年出现的彗星全都是同一颗，并且预言它还会于1758年再来。他于1742年去世，当那颗彗星果然如期重现时，他的历史地位遂得以确立。怀着同样的胜利渴望，18世纪后半叶的天文学家也纷纷去寻找彗星。但是有关彗星的一个恼人问题是，当它们刚刚能被辨认时，它们在望远镜的视野里也表现为暗弱的小光斑，正像那使康德心驰神往的星云一样。有许多曾被那些天文学家认为已经逮着的目标，很快又使他们的成名愿望化为泡影，因为在后继的夜晚观察时那些光斑并没有向太阳趋近因而变大变亮，而是总保持着那个样子。天文学家需要一份包含所有已知星云的表，他们就不再会被愚弄而误以为找到了新彗星。第一个够格的星云表由法国天文学家梅西叶(Charles Messier)从1760年苦干到1784年编集而成。最亮的星云，还有星团，都被验明并准确定位。这可不是因为梅西叶认为它们特别重要，而是因为它们使人讨厌，不得不给标上：“别碰，这不是彗星”。这个表原定的作用发挥得很好，梅西叶本人发现了至少15颗彗星(有人说多达21颗)。而当一位关注星云本质的天文学家出现时，梅西叶的表为他的观测提供了极宝贵的起点，他就是威廉·赫歇尔。

赫歇尔1738年生于汉诺威。同父亲一样，他也是汉诺威卫队乐团的乐师，并曾于1756年随团访问英国。一年后，与法国人的一点小麻烦导致汉诺威被占领，赫歇尔遂永久移居英国。在那儿他成了音乐教师、演奏家、作曲家，到1766年当上了巴斯城八角教堂的风琴手。他对天文学起初只是业余时间涉猎，后来越来越热爱，于是决定

不只是像当时大多数天文学家那样观测太阳、月亮和行星，而且还要去研究最暗弱最遥远的天体。这种强烈的愿望驱使他成为制造望远镜的专家。他用自制的反射镜看到了许多前人从没看到过的暗弱天体，也看到了已知天体的许多新现象。他妹妹卡罗琳(Caroline)于1772年来到巴斯。她同样热爱天文学，成了哥哥的助手，两人一起制造望远镜并对天空沿各个方向作系统观测。终于，在1781年的一次天空搜索中，赫歇尔交上了好运。他开始时以为是彗星的一个天体，原来是一颗从未见过的行星，这是在古代就已知道的五大行星之外找到的第一颗。赫歇尔不失时机，建议把这颗新行星命名为“乔治(George)星”，因为当时的国王是乔治三世。尽管这颗行星最后被定名为天王星，国王仍很高兴，于1782年任命赫歇尔为宫廷天文学家[不同于皇家天文学家，后一职位当时由马斯基林(Nevil Maskelyne)神父担任]。赫歇尔还在同一年成为皇家学会会员，并从巴斯来到温莎，然后又到了斯劳，在43岁那年成了由国王资助的职业天文学家。

从此以后，赫歇尔再不愁没钱建造更大更好的望远镜来探测夜空了。他得到一份梅西叶的星云表，并同卡罗琳一起着手搜寻更多的星云。1802年他俩编出了列有2000个星云的新表，1820年又发表了一份更长的表，那时的赫歇尔已经80多岁，此前在1816年他被封为爵士。赫歇尔对宇宙岛思想的发展功绩甚伟。他应该是直接从赖特的书中获悉这一思想的，因为他所保存的一本赖特的书至今还在，书上还有他在1781年后的某个日子留下的手迹。大约到1785年时，赫歇尔已经确信所有的星云如康德所推测的那样都由恒星组成，而且他和卡罗琳已经用自制的望远镜成功地将许多星云分解成一群恒星。

为了解释恒星为什么会以这种方式聚集在一起，赫歇尔提出了一种宇宙演化的理论，认为弥散的恒星会在引力作用下缓慢地聚集成

群。这是用变化来描述宇宙的最早尝试之一。尽管这个理论的具体内容现在看来并不正确，但赫歇尔设想宇宙本身随时间改变和演化的胆识的确值得称赞。

不幸的是，赫歇尔自己后来没有把关于宇宙岛假说的工作再做下去，因为他注意到有些星云即使用他最好的望远镜也不能被分解成恒星。这一事实为另一种对立的理论提供了证据，该理论认为星云其实只不过是自己发光的物质云。赫歇尔还发现有些不可分解的星云与单个恒星相联系，中间一颗恒星，周围是一团云。这看来又支持了这样的观点，即星云可能是正在形成中的行星系统，恒星和行星正由气体云收缩、凝聚而成。由于赫歇尔的巨大声望，星云是气体云的理论在19世纪初期得到广泛传播，而与之竞争的星云是恒星组成的宇宙岛的理论走了下坡路。天文学家又用了100年才完全搞清楚，原来有两类不同的星云，一类是在银河系内的发光气体云，另一类则是宇宙岛，是远在银河系之外的其他星系。

19世纪后半叶，更大的望远镜和新的观测技术开始使情况变得明朗。威廉·帕森斯(William Parsons)，即第三代罗斯(Rosse)伯爵，出生于1800年的爱尔兰政治家、工程师和天文学家，决心继续赫歇尔中途停止的星云观测。由于赫歇尔没有留下如何制造望远镜的任何说明，罗斯不得不自己重新摸索磨制和安装大型反射镜的技术。最终他造出了一架大望远镜，重4吨，主镜口径72英寸(约1.83米)，镜筒长50英尺(约15米)以上，配以链条和滑轮系统，安装在两根石柱上。有了这个绰号为“科克城巨兽”的有力武器，罗斯和他的助手们到1848年时已经将50个星云分解成恒星，并且注意到其中一些星云呈螺旋形，就像水中的旋涡。这些发现使关于宇宙岛的思想得以复活。几年后，终于得到了有着两类星云的第一个过硬证据。

出生于1824年的英国天文学家哈金斯(William Huggins)，是将光谱技术应用于天文学的先锋。光谱仪是这样一种仪器，它能将光分解成单色成分，并按波长排开，就像彩虹那样。这样得到的光谱上就显示出各种波长的或明或暗的谱线。发光材料(比如某些黄色的路灯)中的不同元素会各自产生不同波长的特征谱线，例如，钠元素会产生一对很亮的黄色线，别的元素产生的谱线不会有与之完全相同的波长，因而在光谱上不会占有同样的位置。于是，光谱学提供了一种通过证认太阳和恒星光谱中的谱线来确定这些天体中所含元素的途径。继基尔霍夫(Gustav Kirchhoff)和本森(Robert Bensen)于19世纪50年代后期对太阳光谱的研究之后，哈金斯也于该世纪60年代投入进来。他和一位朋友化学家米勒(W. A. Miller)一起查明，恒星光包含的谱线与太阳光的谱线一样。这就是说，恒星的化学成分与太阳的基本相同。

但是，当哈金斯把目光投向星云时，不同的情形出现了。有些星云，例如猎户座的那一个，还有以形状而得名的蟹状星云，它们发出的光在光谱仪中并不显示预期的谱线式样，而像是热气体云所发的光。后来的类似研究又表明，其他星云，包括那些旋涡状星云的光谱线的确与恒星的光谱相似。

难题的各个片断已开始汇集，但是尚未组合成形。康德的猜想提出来不过一个世纪，它毕竟只是空想学者的作品。气体云显然是银河系的成员，其他那些星云即使由恒星组成，它们很可能也是银河系的成员，或者是小的恒星系统，或者是正在形成中的恒星，而不是像银河系那么大的整个星系。但问题是，银河系究竟有多大呢？19世纪末对银河系尺度的认识还非常粗略，部分地是通过估计恒星必须有多远才会如其呈现的那样暗弱。流行的观点仍然是，银河系就是宇宙，

虽然它好像有着康德所提出的那种扁平盘状结构。太阳和太阳系被认为处于靠近盘中心的地方，而那些已使天文学家饱受困扰的旋涡状和椭圆状的星云，则被认为也在银河系内。实际上，截至那时为止，天文学家对宇宙大尺度上的观测已超过100年，尽管他们自己并未意识到。我们现在知道，星云是构成宇宙的单元。但是，为了认识它们的真实本质，必须做一件超越以前的简单观测的事。天文学家在真正领会自己观测到的现象的意义之前，必须先量度宇宙，先掌握银河系之内和之外的距离尺度。

注　释：

1. 《星云世界》(*The Realm of the Nebulae*)，pp. 23 & 24。文中提及的该书的细节在参考文献中给出。

第二章　天有多高

哈雷是有记录的第一位意识到恒星在运动的天文学家。既然恒星之间有相对运动，它们就不可能全都固着在一个围绕地球的大球的内表面上。有着不同亮度的恒星在运动着的证据，也就是它们散布在三维空间、与我们有着各不相同距离的证据。哈雷于 18 世纪初作出的这一发现，是否定恒星固着在一个比土星轨道(早先天王星尚未被发现)稍大的球面上的图像的第一个直接观测证据。这一发现为 18 世纪后期的思想家，如赖特和康德，作出关于银河本质的推测铺设了道路，它还以一种直接然而缓慢的方式导致了对宇宙尺度的认识。

哈雷的宇宙

哈雷在作出上述发现时已经是一位受尊敬的、硕果累累的资深天文学家。他出生于 1656 年，在牛津大学读书时就撰写并出版了一本关于开普勒所发现的行星绕日轨道定律的书。开普勒定律为牛顿关于引力本质的研究提供了关键线索，而大学生哈雷的书则受到了当时的

皇家天文学家(是第一任，实际上称为“天文观测家”)弗拉姆斯蒂德(John Flamsteed)的注意。当哈雷没有完成学业而离开牛津时，弗拉姆斯蒂德帮助他得到了一份从事天文研究的工作。他被派到南大西洋的圣海伦娜岛干了两年，任务是绘制南半球天空的星图。他于1678年返回英国，并立即入选皇家学会，当时年仅22岁。但是他并没有走纯学术的道路。

哈雷在此后30年中的传奇经历包括：周游欧洲以会见其他科学家和天文学家，当过两三年设在切斯特的制币厂的代理主管、一段时间的皇家海军战舰“帕拉蒙号”的指挥官、还有代表政府驻维也纳的外交官。与此同时，他还对认识磁、风、潮汐作出了重大贡献，并且对牛顿《原理》一书之完整出版起了关键作用——是他说服了牛顿同意这样做，他甚至出资赞助。1703年他成为牛津大学的几何学教授(这对一个中途退学的学生来说真够不错了)，1720年他继弗拉姆斯蒂德之后被聘为皇家天文学家，并担任此职直到1742年85岁时去世。他对认识恒星的贡献是在18世纪的头20年里作出的。

哈雷始终对古代的天文学文献很有兴趣，并且翻译过一些希腊文著作。他在1710年开始研究托勒玫(Ptolemy)写于公元2世纪的著作，由于自己曾经做过编制星表的工作，他特别注意托勒玫的恒星位置图。“托勒玫”星图实际上比托勒玫本人还要早，它可以上溯到公元前3世纪依巴谷(Hipparchus)的工作。这是第一张重要的星图，标有800多颗恒星的位置。托勒玫为后世保存下了这张图，并且自己标记了更多的恒星位置，使图中恒星超过千颗。那些恒星位置大多与哈雷及其同时代人所作的观测符合得很好。但是哈雷在1718年注意到，有三颗星，即天狼星、南河三和大角星，并不在依巴谷和托勒玫所看到的位置上。差异是如此之大，看来不像是古希腊人搞错了。再

说，他们怎么会对上千颗恒星的位置作了精确观测而偏偏要搞错三颗呢？例如，大角星在1718年的视位置与托勒玫的记录相差为满月宽度的两倍，即整整1度。哈雷认为，这意味着大角星，还有其他恒星，已经在过去的许多世纪里从希腊人所记录的位置上移开了。这种移动太慢，一个人在有生之年不可能用肉眼看到，但是经过几代人的时间就可以显示出来。

恒星的这种移动现在被称为自行。首先被证认出存在自行的这三颗星属于天空中最亮的八颗恒星之列。对这种“巧合”的自然解释是，有的恒星显得明亮是因为它们比别的恒星离我们近得多，而我们能看出它们在几个世纪里的运动也是因为它们离得近。正如高空中的飞机看上去在缓慢移动，而几米以外一个孩子却在骑着自行车疾驶而过一样，离得近的恒星应该比离得远的表现出更大的表观运动——如果它们都以大致相同的速度在空中运动的话。绝大多数恒星都离得太远，以至于从地球上用几千年也看不出它们有表观运动。只有少数几颗离得近的恒星才会被识别出在看似“固定”的恒星背景上运动。

我们不妨暂停一下，先来看看今天的天文学家所测量出的恒星自行的大小。月亮在天空中的张角大约是半度，或30角分[1]。正如每1度分为60角分那样，每1角分又分为60角秒。望远镜观测得出，木星在最靠近地球因而显得最亮时的张角只有50角秒，因而用肉眼看去仍然只是光点而不是圆盘。已观测到自行最大的恒星名叫巴纳德星，因被美国天文学家巴纳德(Edward Barnard)于1916年发现而得名。这颗星看上去很不明显，实际上也是又小又暗，所以以前没有被注意到。但是它离我们是如此之近，以至于以每年10.3角秒的创纪录速度在空中飞驰(也就是说，它在5年中越过的角距离相当于木星最亮时的张角)。其他恒星被测得的自行一般都小于每年1角秒。为

了把这种微小的自行变换成在空中运动的速度，天文学家必须找到估测恒星距离的途径，而他们首先又必须作出对太阳系尺度的精确测量。

从地球到太阳

测量宇宙尺度的第一步所采用的方法，就是绘制地图时在地面上测量位置和距离所用的三角法。天文学家通常使用视差这个术语，你只要看看自己的手指，就能明白那是怎么回事。

伸直手臂，竖起一根手指，闭上一只眼。先记住你的手指落在房间墙壁背景上的位置。再闭上你原来睁开的眼，睁开你原来闭着的眼。手指似乎在背景上移动了，跳到了一个不同的位置。这是因为你的两只眼睛是从不同的方向来看手指的。现在把手指放到鼻子跟前，并再次换眼睛看。手指位置的变动，也就是视差，比上次要显著得多。物体离得越近，其视差效应就越大，所以通过测量视差就有可能算出物体的距离。如果你分别从两个相距甚远的地点去观察同一个遥远的物体，你就能用这种效应来确定其距离。这个方法对一座山或月亮或太阳系内的行星都同样适用，只要你能够从相距足够远的两个地点去进行观测，也就是说能为你的三角形找到一条足够长的基线。例如，为了测出月亮的距离，所需做的一切就是在同一时刻从两个相去甚远的观测站各自记下月亮在遥远恒星背景上的位置。只要两个观测站之间的距离已知(当然要考虑到地面的弯曲)，就能用上述观测结果构造出一个以月亮为顶点的虚拟三角形，于是运用简单的几何知识就能算出三角形的高，也就是月亮到地球的距离。

上述方法对月亮适用，是因为它离我们很近，只有大约400 000千米，视差效应也就相当显著，虽然虚构的三角形又高又瘦，其角度还

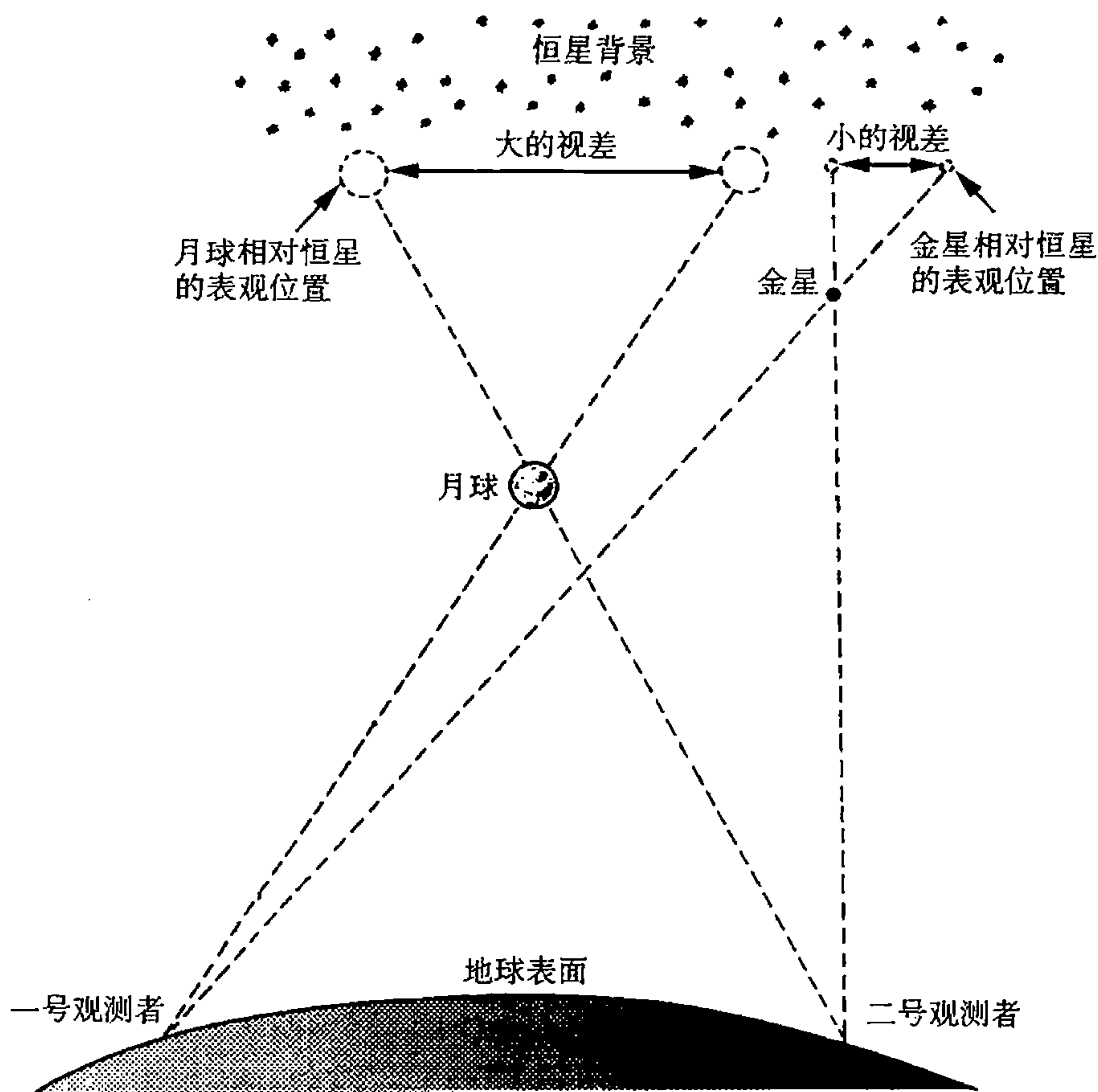


图 2.1　视差，即遥远物体由于对其进行观测的位置改变而表现出的移动，可以用于测量太阳系内的距离。

是能很容易地测出来。[2] 当天文学家试图用同样方法测量行星的距离时，事情变得稍稍麻烦一点。视差效应很小，即使从地球的两端去观测，那个以地球直径为基线、以（比如说）火星为顶点的三角形仍然瘦得难以置信。尽管如此，在适当条件下角度还是能被测出来，距离也就能算得出来。

在望远镜发明之前对行星的视差测量是不可能的，即使有了望远镜，天文学家也不得不远征到地球的遥远角落以获得足够长的三角形基线。对火星视差第一次真正成功的测量是由一帮法国人于 1671 年作出的。里歇尔（Jean Richer）领队远赴法属圭亚那的卡宴，而卡西尼（Giovanni Cassini），一位意大利出生的法国人，则留在巴黎进行观

测。双方在预定的同一时间各自记下了火星在天空中的位置。里歇尔返回后，他们比较了相互的记录，并推算出了地球到火星的距离。有了这个结果，他们还进一步利用开普勒定律来计算火星到太阳的距离和地球到太阳的距离。他们得出的日地距离是 1.4 亿千米，仅比现代测量值小 1000 万千米。那么，在 17 世纪和 18 世纪，视差方法又是如何得到改进的呢?

当哈雷孤独地在圣海伦娜岛值班时，他在编制星表之余还有充分的时间思考。他观察并记录了一次水星凌日，也就是水星表现为一个小黑点缓慢地越过明亮的太阳圆面，这是一种相当罕见的现象。他意识到，这种现象提供了用三角法测量太阳系的又一途径。由于视差效应，水星看上去开始触及太阳圆面的精确时间取决于你在地球上的什么地方进行观测。哈雷时代的天文学家知道在 1761 年将有一次金星凌日，这是比水星凌日更为罕见的现象。哈雷在自己的手稿里详细论述了如何最佳地利用在世界各地观测这次凌日的资料来推算地球到金星和到太阳的距离，并于 1716 年发表了这些手稿。虽然他在那次金星凌日之前 19 年就去世了，但他的影响仍是促成天文学家同心协力测量金星视差的因素之一。62 个观测站跟踪了 1761 年的凌日，对 1769 年的凌日也作了类似的努力。在所有资料被汇集处理之后，算出的日地距离为 153 000 000 千米，与现代测量值 149 600 000 千米很接近。尽管在后继的年代里测量精度有了提高，但在总体图景的意义上讲，18 世纪末的天文学家已经了解了太阳系的尺度。他们已经测出了地球到太阳的距离，这使他们获得了巨大的新机会；因为这一距离可以作为一条新基线用于三角法和视差测量。日地距离在天文学中是太重要了，因而被称为天文单位，记作 AU。有了这条 1.5 亿千米长的基线，也许就有可能去测量地球到恒星的距离了。

从太阳到恒星

1761 年金星凌日之后又过了 3/4 个世纪，天文学家终于着手来测量几颗邻近恒星的视差。所依据的原理是很简单的。既然地球轨道的半径是 1.5 亿千米，相隔 6 个月的两次观测就是在太阳的相对两侧，亦即在一条长达 3 亿千米(2AU)的基线的两端进行的(从得到对恒星距离的基本估计的角度看，基线的长度是 2.9 亿千米还是 3.1 亿千米都没关系，结果都会大致正确)。剩下的事情就是做些简单的几何计算，来得出恒星应该有多远才会显示出相应于这条基线的视差位移。事实上，天文学家还据此原理定义了一个新的长度单位。如果从一条长度等于日地距离的基线的两个端点看去，一颗恒星显示出 1 角秒的视差位移，则它到该基线的距离就称为 1 秒差距，记作 pc。也就是说，若以长为 3 亿千米的地球轨道直径作为基线，一颗距离为 1 秒差距(1pc)的恒星将显示出 2 角秒的视差。

1 秒差距是日地距离(即天文单位)的将近 206 265 倍，即略大于 30 万亿(3×10^{13})千米。速度为每秒 30 万千米的光要用 3.26 年才能跑完 1 秒差距的距离。但是，没有任何一颗恒星与我们如此靠近，以至于能显示出这么大的视差。这也就是为什么直到 19 世纪 30 年代才第一次测出恒星视差的缘故。

如果你试图测量小于 1 角秒的恒星位置变化，很显然你需要一张至少以同样的精度标出恒星位置的星表。18 世纪初期最好的星表是由弗拉姆斯蒂德编制并在他死后于 1725 年出版的。它所给出的恒星位置精确到大约 10 角秒，这在当时已是惊人的成就，但仍不足以用于测量恒星位置的变化。第三任皇家天文学家即哈雷的继任者布拉德利(James Bradley，生于 1693 年)作了巨大的努力，试图测量一颗名为天龙座 γ 的恒星的视差。他并没有成功，但在这个过程中他改进

了观测技术，造出了更好的仪器，并且使天文学家加深了对自己所作观测的理解。布拉德利发现，天龙座 γ 星在天空中的位置的确在一年中有所移动，但这不像是视差，因为他注意到所有恒星都有同样的现象。他终于意识到这是由于地球绕太阳运动造成的。来自遥远恒星的光线由于地球的运动而显得倾斜，这与当你行走时觉得本来垂直下落的雨滴迎面扑来是一样的道理。这种星光倾斜效应在地球轨道上的不同地点(亦即一年中的不同时间)看去是不同的，因为地球的运动方向在不断改变。由于光速很大，这种效应也就很小，但在测量视差所要求的精度水平上仍可觉察出来。

布拉德利把上述效应称作“光行差”，它使恒星的视位置在一年中发生 20.5 角秒的移动。这一发现证实了光具有有限的速度(尽管这在 18 世纪晚期已不是什么新闻)，给出了一个接近于现代值的光速估算值，并且证实了地球的确在空中运动。但是布拉德利在测量恒星位置时仍不得不考虑更多的干扰，其中包括被他称为章动的地球晃动，后者乃是因地球并非严格的球形所致。所有这些努力的成果是一份新的星表，以前所未有的精度记载了大约 3000 颗恒星的位置。这份表分为两部分，于 1798 年和 1805 年先后出版，那已是布拉德利去世 30 多年之后的事了。

许多天文学家采用了布拉德利的技术，其中最著名的一位是 1784 年出生的德国人贝塞尔(Friedrich Bessel)。他列出了 30 000 颗恒星的位置，并且是大约同时(即 19 世纪 30 年代末)各自独立地解开视差难题的三位天文学家之一。赫歇尔也是试图测量恒星视差但未能如愿的天文学家之一。他尝试了一个巧妙方案，并不需要星表绝对精确，那就是观察天空中相互很靠近的一对恒星，如果它们的靠近只是视线方向上的重合，实际上其中一颗离我们很远而另一颗很近，则近

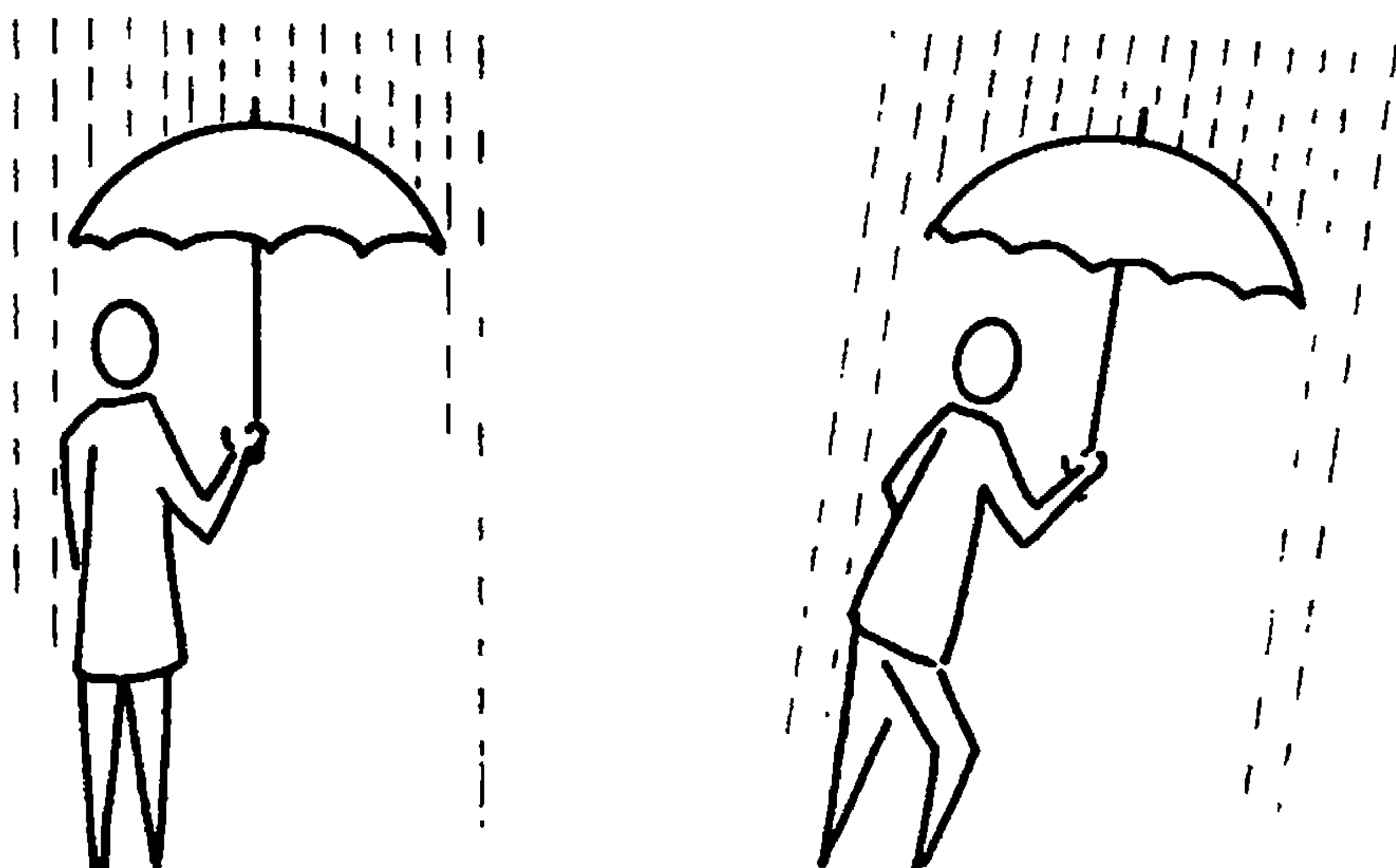


图 2.2 当你行走时觉得空中垂直下落的雨滴仿佛在迎面扑来。来自遥远恒星的光线由于地球在空中的运动也以同样的方式倾斜，这种效应叫做光行差。

的那一颗就会显示出视差位移。较近恒星的这种移动只需相对于较远那颗恒星来测量，而不需要任何绝对标准。但是，赫歇尔发现他观看的那些成对的恒星都是真正成对，真的靠得很近，像月亮和地球那样相互围绕旋转。这其实是一项重要的发现，但并不是他原来所期望的。

突破的来临必须等到观测已足以显示恒星位置的微小移动，理论也足以扣除所有其他因素的影响，诸如光行差和章动，它们也会导致恒星位置随季节的变化。成功是在时机成熟时到来，而不是在此之前；但当时机的确成熟时，成功又是突然降临的。19 世纪 30 年代三路攻势的发动者之一是贝塞尔，他挑选了天鹅座 61，因为那颗星有很大的自行，每年 5.2 角秒，因而必定离得很近；第二位是 1798 年出生的苏格兰人亨德森(Thomas Henderson)，他在南非进行研究，选的是半人马座 α，那是夜空中第三亮的恒星，因而也必定离我们很近；还有一位是斯特鲁维(Friedrich von Struve)，生于 1793 年，是个在俄国工作的德国人，他看中了织女星(又名天琴座 α)，夜空中第四明亮的

恒星，理由自然与亨德森的一样。贝塞尔首先于1838年末宣布了他的成功；亨德森实际上最先完成决定性的观测，但直到1839年1月返回英国时才公布自己的发现；而斯特鲁维的结果就像在蛋糕上加糖，于1840年面世。他们测得的三个视差的确都很小，天鹅座61是0.3136角秒，天琴座α是0.2613角秒，半人马座α则是1角秒(后来得到的更精确值是0.76角秒)。半人马座α的视差是已知最大的，这颗星(实际上是三颗相互绕转的恒星组成的系统)是太阳最邻近的伙伴，与我们相距1.3秒差距(4.3光年)。天琴座α即织女星，离我们8.3秒差距(27光年)；现在知道天鹅座61是一对双星，距离为3.4秒差距，即约11光年。天文学家第一次真正领会到我们太阳系是多么孤独，我们周围的空间是多么黑暗而空旷。太阳到最邻近恒星的距离是它到今天所知最遥远的行星冥王星距离的7000倍。*一旦知道了哪怕是寥寥几颗恒星的距离，天文学家就能算出它们的真实亮度，并从而大致估计出那些太遥远而无法测出其视差的暗弱恒星的距离。运用这一技巧并加上其他工具，天文学家终于在19世纪后半叶开始从定量的意义上了解我们银河系的尺度和形状。但他们只是在20世纪才首次得以把视差方法发展到能用于大量恒星，并随之进军到星云世界。

表2.1　到一些近邻恒星的距离

	距　　离	
恒　　星	光　　年	秒　差　距
半人马座 α	4.29	1.32
巴纳德星	5.97	1.84
沃尔夫359	7.74	2.38
天狼星	8.7	2.67
天鹅座61	11.1	3.42
小犬座 α	11.3	3.48

* 自2006年8月起，冥王星已被降为“矮行星”。今最遥远的行星是海王星。——译者

通向银河系的台阶

使视差方法得到重大改进又用了 60 年时间，那是在望远镜末端用照相底片代替肉眼作为标准的观测手段之际。照片比起肉眼有两个关键的优点。首先当然是它能给出直接的、可长久保存的恒星位置记录，于是就可以对之作从容研究和精细测量，甚至用显微镜来精确测定星像的相互位置。其次，照相底片能“看见”非常暗的物体，远非人眼所能及。底片曝光的时间越长，落到它上面的光就越多，暗弱的像就会变得越明显；而你用眼睛一开始就看不见的东西，不管你盯多久还是看不见。所以，天文照相术提供了多得多的恒星用于研究，并使更精确地测量其中每一颗星的位置成为可能。1900 年，当照相术开始被采用时，测得视差的恒星只有 60 颗。半个世纪后即 1950 年，已经知道距离的恒星数目接近 1 万，不过并不是所有这些恒星的距离都直接得自视差测量。

特别是有三项技术成了天文学家的台阶，使他们能从太阳周围的狭小空间——由视差测量可靠地指明尺度的范围只有 30 秒差距或者说 100 光年——攀升到整个银河系。恒星并不是都有完全一样的亮度，这一点很快就清楚了。继哈金斯的开创性工作之后的恒星光谱研究则进一步表明，恒星是分族的，同族成员性质相似。如果一颗恒星光谱中特定的谱线样式已记录在案，并且已通过视差测定其距离，从而真实亮度也已得知，后来又发现另一颗更遥远的恒星有着同样的光谱类型，那么就可以合理地猜想后者也有着同样的真实亮度。当然，后面这颗星看上去更暗，通过精确地测量它的视亮度（或暗弱的程度），并与前面那颗距离已知的恒星作比较，就可以估算出它的距离。

另外两项技术都出于几何学上的妙招，也都有赖于光谱学。光谱

学中一个至关重要的效应是谱线移动，就是当光源做趋近或远离观测者的运动时，观测者接收到的光谱中的特征谱线位置与光源静止时的谱线位置相比会发生移动，在本书下文关于河外星系的故事里将看到这种效应更大的重要性。首先设想一个物体在远离我们而去。它发出的波，不论是恒星发出的光波，还是地面上的物体（比如一辆警车）所发出的声波，就会由于波源的运动而伸展，也就是波长变长。对声波而言，它使声音听起来变得低沉，对光波而言，则是使可见光的谱线朝光谱中红色一端移动。[3] 反之，当波源朝向我们运动时，波被挤压，波长变短，于是声音变得尖锐，或是光谱线朝蓝色一端（或者说紫色一端）移动。

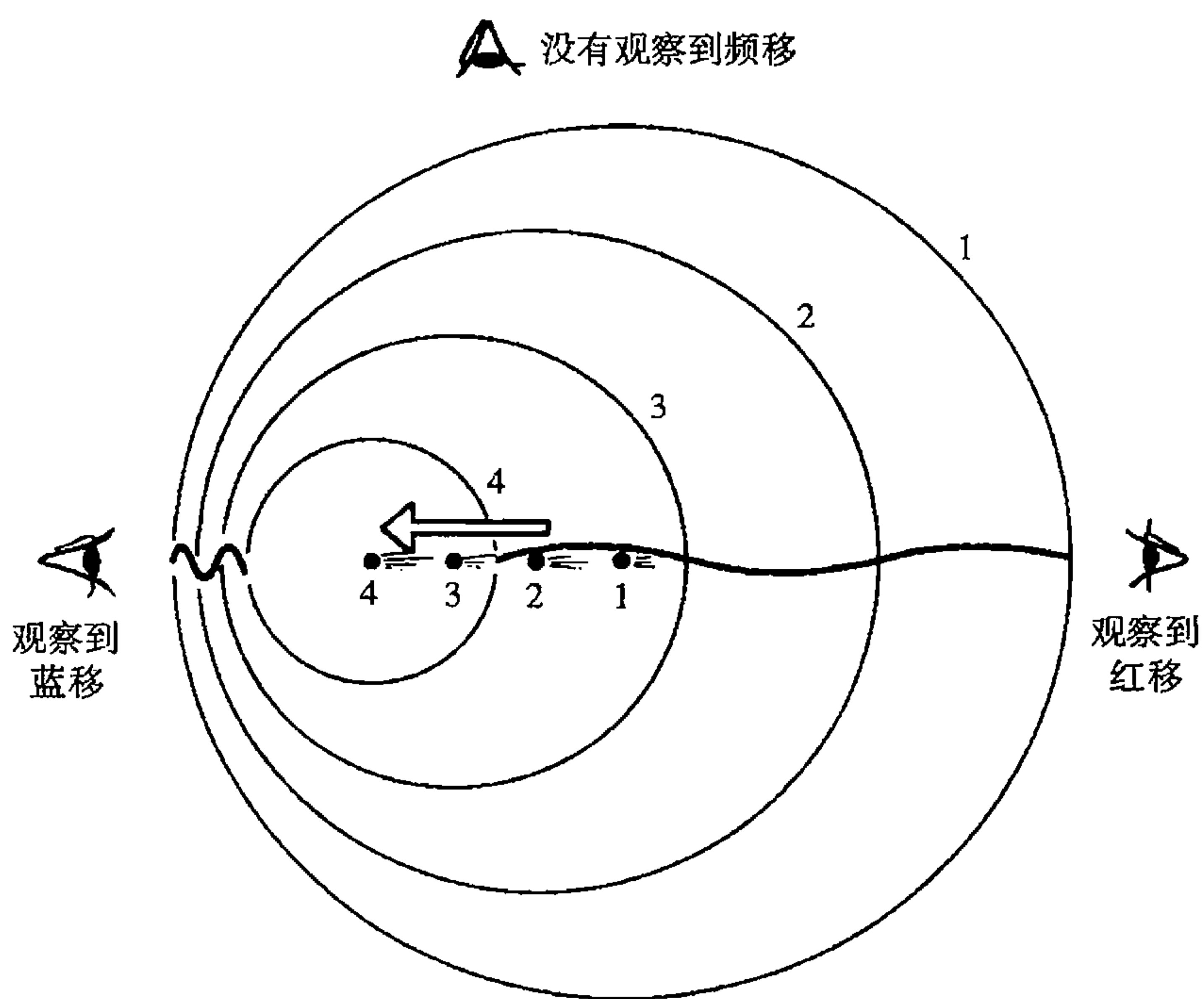


图 2.3　多普勒效应挤压由趋近观测者的物体发出的光波，拉伸由退行物体发出的光波。圆圈表示该物体在各个标号地点发出的光波。虽然光在所有方向都以同样速度传播，但是，在光源运动方向上圆圈却被挤拢了。

声波的接收频率依赖于声源相对于接收者的运动速度这一效应是由奥地利物理学家多普勒（Christian Doppler）于 1842 年发现的，因而

被称为多普勒效应。多普勒本人还意识到，来自运动光源的光也会有类似变化。法国物理学家菲佐(Armand Fizeau)于1848年首先对光的红移或蓝移效应作出了清楚的描述。

重要的是，谱线移动的多少是取决于恒星趋近或离开我们的速度大小。由于整个光谱全都受到挤压或拉伸，那些特征谱线，例如钠线，所出现的波长会向红端或蓝端移动，移动量取决于恒星沿视线方向的运动速度。于是，通过测量那些熟悉的谱线在恒星光谱中出现的精确位置，并与同种谱线在地球上实验室光源的光谱中的波长作比较，天文学家就能推断该恒星是在朝向还是背离我们运动，是以多大速度运动。当然，这样能得知的只是沿视线方向的运动。恒星也可以有与视向垂直的运动，即有横向速度。恒星在空中的实际运动速度可以通过把上述两种由观测得知的速度，即横向速度或者说自行运动速度和由光谱线的红移或蓝移确定的视向速度，用几何法合成而得到，实际运动方向会与视向成某种交角。

那么，怎样把测得的速度与距离联系起来呢？有一种方案仅仅适用于星团。星团是在空中一起运动的一群恒星，离太阳不算太远。一群全都朝同一方向运动的恒星，实际上是在沿着许多条像铁轨那样的平行线奔跑。正如铁轨看似汇合在远处的一点那样，假如该星团离我们比较近，那么这群恒星的运动(由多年观测定出的自行来量度)便亦是如此，即仿佛会在空中某一点汇合，这就有一个很大的好处，即可以知道星团在空中的运动是朝哪个方向(与视线方向成多大交角)。于是，当天文学家测出星团中恒星的多普勒运动时，他们就不仅能得出恒星沿视向的速度，还能由恒星真实运动方向与视向的交角来进一步推知总速度。总速度的另一个分量当然就是恒星真实的(以千米每秒为单位的)横向速度。既然早已得知恒星的自行(以角秒来量度的横越

天空的运动），于是就可以虚构一个非常瘦长的三角形来推算出那个星团必定有多远，才能使以千米每秒为单位的那个真实横向速度表现为在天空中以角秒每年为单位的这个移动。瞧，多么美妙的技巧，尽管它只对与太阳的距离在几十秒差距以内的星团才有效，它仍让天文学家得以求出了许多星团的距离。其中有一个特别值得一提，那就是毕星团，由大量不同光谱类型的恒星组成，坐落在约40秒差距处。距离有了，各类型恒星的真实亮度就可以确定，于是就可以用上述第一项技术来估算因太遥远而无法测出自行的同族恒星的距离。

另一项重要的几何技术听起来似乎太天真，其实却很有效。取一大批恒星，只要它们离我们足够近，因而能测出自行运动。从地球上看去它们可能是在同一个大方向上，也可能是散布在全天空，只是由于有同样的颜色或同类型光谱而被选中。它们的运动有的朝东有的朝西，有的快有的慢。但是银河系作为整体，至少是在太阳周围区域，看来既没有自行坍缩，也没有向外爆炸。总的说来恒星是有秩序的。所以，所有那些无规运动平均起来必定互相抵消。不妨猜想，平均而言，一颗恒星往东和往西的机会是一样大的。那么，如果我们把所有入选恒星的速度都沿视线方向加起来并且取平均多普勒速度，我们应该可以预期这批恒星在其他任何方向上，尤其是在横向上，也有同样大的平均速度。假定这是对的，就可以通过把预期的实际横向速度与测量的自行角速度对照，而赋予这批恒星一个“平均距离”。

这一技术叫做“统计视差”。如果只用在一颗恒星上它的确起不了什么作用，但是所取的恒星越多，所作的平均就越可靠，所以用它来指示某些恒星的距离并不太坏。特别重要的是，已经证明可以用这一技巧来大致估计如下一群恒星的“平均距离”，在这群恒星中有那么几颗是特别类型的“造父变星”。造父变星是测量整个银河系乃至

宇宙的标尺。如我即将讲到的，我们对宇宙的测量依赖于掌握少数造父变星的距离。现在已有其他的技术(我不准备在此细述)将以前对造父变星距离的初步估计作了改进。那些技术涉及的是恒星的颜色和视亮度。取一群在一个确有物理联系的星团中的恒星，其中各个恒星的颜色可以标在一幅所谓的颜色—星等图上。这群星在图上组成的一条线的位置取决于该星团从地球上看去的亮度，而且可以调节这个位置，以使所有这样的星团都重合在一条标准线上，各个星团的线与标准线的差距就反映了各星团距离的不同。换句话说，假定每个星团中的恒星都服从同样的物理原理(若非如此我们就根本不必去搞什么天文学了)，我们就能通过给每个星团设定一个距离而把它放在颜色—星等图的适当位置上。但是我们仍然必须首先由某种视差技术来知道至少某一个星团的距离，以标定颜色—星等图上的距离尺度。

当然，天文学家也能看到并研究比他们能由视差来测定距离的更加遥远得多的天体。但是他们只能猜测那些更遥远的天体究竟有多远，他们也只能估计银河系的范围。关于在银河系之外还存在其他星系的思想对许多天文学家来说显得模糊乃至荒谬就并不奇怪了，一直到出现一把新的测量标尺才使局面有所改观，这把尺子能够伸过广漠太空，对一些河外星云的距离作出直接测量。

造父变星标尺

正如从太阳系走向银河系的第一步是靠着测出最邻近恒星的距离那样，从银河系走向整个宇宙的第一步是靠着测出我们在河外空间的最近邻居的距离。首先结识的芳邻是两个叫做麦哲伦云的星云，在南半球的天空用肉眼可见。它们之得名是由于探险家麦哲伦(Magellan)在1521年的环球航行时作了记录，那是欧洲人第一次知道这两块

云。它们一大一小，看上去像是银河被折断时所形成的碎片，当然在16世纪时没有人知道它们和银河本身究竟都是什么。它们一直被天文学家所冷落，直到威廉·赫歇尔之子，约翰·赫歇尔(John Herschel)，于19世纪30年代对南天恒星和星云作普查时才被重新记起。到20世纪初时已没有人怀疑这两个星云和银河一样也是恒星的集合。但是，那时候星云是宇宙岛的思想并不合时尚，天文学家普遍认为麦哲伦云是银河系的组成部分，或者是恰好处在银河系之外的很小的半独立系统，即由银河系引力控制的伴星系。对麦哲伦云乃至宇宙尺度的正确认识，不是来自蓦然间的灵机一动，也不是由于观测到了什么新现象，而是由爱德华·皮克林(Edward Pickering)于19世纪后期开始在哈佛学院天文台对数十万颗恒星进行编表和细致分析的结果。

皮克林1846年出生于美国马萨诸塞州的波士顿，19世纪60年代和70年代在刚成立的麻省理工学院教物理学，1876年被聘为天文学教授和哈佛天文台台长。此后40年中他负责编集几个新的星表，每一个都比前一个更大更好。他更是整整一代天文学家的灵魂。与当时社会上常见的一样，仔细收集整理恒星位置和亮度的资料并用黑墨水笔工整地填入极长的表格的枯燥工作，交给了可以少付工资的妇女；而一反世俗的是，皮克林允许并鼓励那些妇女中的少数人去做更高级的工作，从而使她们在当时几乎清一色由男人占据的学术界获得了一席之地。那些幸运者之一是勒维特(Henrietta Swan Leavitt)，她被分派的任务是从南天照片中证认变星，照片是由爱德华·皮克林的兄弟威廉·皮克林(William Pickering)在秘鲁的一个观测站拍摄的。

勒维特生于1868年，就读于女子高等教育学校(即后来的拉德克利夫学院)，1895年作为志愿研究助理加入皮克林在哈佛大学天文台

的队伍，1902 年获得永久职位，很快又成为一个研究室的负责人。皮克林无疑乐意他的队伍中有人以自己的经验、耐心和能力从来自秘鲁的大堆照片中找出有意义的东西，但是在 1895 年时没有谁看到有关勒维特此后 17 年里所获成果的任何迹象。

变星，即亮度有变化的恒星，显然为天文学家所关注。大多数恒星似乎总是那样子，至少在一个人的有生之年是如此，而任何违反常规的东西总是注意的焦点。有些变星实际上是两颗互相绕转的恒星，两者轮番遮挡(即掩食)对方。另外一些，如我们今天知道的，是在作脉动，自我膨胀及收缩，很有规则地不断重复着，于是它们的光就时强时弱。造父变星就属于这一类。还有很少数恒星有剧烈变化，在度过平静的常规生涯后向外爆炸，造成短暂而猛烈的能量涌泻，然后快速收缩并暗淡下去，留下一具残骸。天文摄影术的巨大优越性之一，正是通过比较相隔数日、数月或数年所拍摄的照片，就有可能证认出所有这些类型的恒星活动。你甚至可以研究那些在拍摄时并不知道有什么重要性的现象。勒维特一生共证认出了 2400 颗变星(这是截至她 1921 年去世时已知变星总数的一半)，还有 4 颗爆发的恒星，即所谓新星。正是她在对一种特别类型变星的研究中找到了打开宇宙之门的钥匙。

造父变星家族得名于仙王座 δ *，1784 年由年轻的英国天文学家古德里克(John Goodricke)证认的一颗变星。古德里克在此后两年去世，年仅 21 岁。造父变星的亮度呈现特有的很规则的变化，但是不同的造父变星有着不同的变化周期，有的短于两天，有的长过百日，平均大约是 5 天。它们之所以被归为一族，是由于其特有的明暗变化

---

* 这颗星在中国古代称为“造父一”。——译者

方式和相似的光谱。这里显然有一个有趣的问题，既然每一颗造父变星都有其恒定的光变周期，为什么不同造父变星的周期又各不相同呢？勒维特在从照片上证认造父变星(和其他变星)并找出每颗变星的光变周期和平均亮度的艰苦工作中，逐渐获得了这样的印象：越明亮的造父变星，其亮度变化的周期越长。

当勒维特于1908年发表关于自己工作进展的初步报告时，她就讲了这么多。又过了4年即到1912年，造父变星的这条规律才确定下来，给银河系确立一个精确的距离尺度也随之真正有了希望，而这一切都要归功于麦哲伦云。

勒维特那时已经在两个麦哲伦云较小的那一个里证认出25颗造父变星。它们很清楚地显示出亮度与周期之间的关系，而对银河系里的造父变星就几乎看不出这种关系。其中道理不难明白。银河系里的恒星散布在与我们各不相同的距离上，有的很靠近，有的却远了十倍、百倍、甚至更多。如果一颗恒星的真实亮度和距离都是另一颗的两倍，则前者看上去较暗，因为视亮度与真实发光本领(即光度)成正比而与距离的平方成反比。所以，造父变星的周期—光度关系被银河系内的距离效应掩盖了。

但对小麦哲伦云里的恒星来说，情形有所不同。这个星云离我们如此之远，以至于其中的所有恒星都可视为大体是在同样的距离上。它们中的这一颗可能比另一颗离地球稍稍近些，但是绝没有一颗的距离能是其他任何一颗的两倍。星云中各恒星与地球距离的差异远远小于星云到地球的平均距离，正如对我这个在英格兰的一个小村庄写书的人来说，纽约的任何人都离得差不多一样远。离我最近的小镇在大约1英里(1.6千米)之外，可小镇的那一头与我的距离是这一头的两倍多。当我打算去小镇的这部分或那部分时，1英里的差别是很重要

的。但是，我可以认为纽约的时代广场和自由女神像与我离得一样远，几英里的差别与北大西洋的宽度相比是微不足道的。

这就是勒维特为什么能用小麦哲伦云的25颗造父变星得出这个家族的光度与周期之间关系的原因。她发现，例如，如果一颗造父变星的周期是3天而另一个是30天，则后者比前者要亮6倍。假定这一来自小麦哲伦云造父变星的规则对所有造父变星都成立，这立即意味着银河系里的造父变星可以用来指示恒星和星团的相对距离。但是没有谁知道哪怕是一个造父变星的真实亮度，所以距离尺度并没有定标。天文学家有了一把测量银河系的尺子，但是并不知道这把尺子的长度。他们能说出一个恒星或星团的距离是另一个的两倍，但不知道究竟是多少千米。所以，他们还是不知道，麦哲伦云究竟是银河系内的小系统，抑或本身就是遥远的星系。

只过了一年就有了答案。赫茨普龙(Ejnar Hertzprung)，一位丹麦天文学家和物理学家(出生于1873年，并且直至1967年去世之前在研究上仍保持活力)，用一种统计视差技术的变体首次估算出了一些较近造父变星的距离。技术虽不完美，却为他指示了一两颗造父变星的实际距离。有了距离，他就能很容易地由视亮度算出真实亮度。于是，要得出其他任何一颗造父变星的实际距离就很简单了，因为由那颗星的周期就可以知道它实际上比已有的样本亮多少或暗多少，也就是知道了它的真实亮度，再由视亮度就可以推算出距离。赫茨普龙断定，小麦哲伦云在30 000光年(约10 000秒差距)之外，这比任何人曾经猜想过的都遥远得多。但是，这个结果并没有马上使天文学家看清宇宙的真实尺度。其原因有二。第一，由于赫茨普龙没有考虑到太空中的尘埃会遮挡遥远恒星的部分光线，因而使恒星看上去比实际上要暗，他的定标稍稍偏低，现在最好的计算结果更为惊人——大麦哲

伦云与我们的距离是 170 000 光年或 52 000 秒差距，小麦哲伦云则是 63 000秒差距。第二，天文学家太忙于用这把奇妙的新尺子去测量银河系的尺度，在随后几年里还来不及去操心银河系外面是什么。越过麦哲伦云并进入真正的宇宙，需要的是想象力的新飞跃和新一代的望远镜。然而，在这两个飞跃实现之前，还是应该首先承认那些天文学家的巨大成就，他们描画出了我们的银河系，他们使用的技术也为后来对更高层次空间的研究打下了基础。

银河系的尺度

两个来自完全不同背景的人一起为迈出认识宇宙尺度的下一步作出了主要贡献。海尔(George Ellery Hale)是 20 世纪甚至有史以来最伟大的望远镜建造者，即使把自伽利略、赫歇尔、罗斯的时代以来的技术进展考虑在内也是如此。他是一个有钱人，是一名电梯制造商的儿子，1868 年生于芝加哥，顺利地沿着常规的教育途径进入麻省理工学院，1892 年受聘为芝加哥大学天文学教授。海尔对天文学的热情是由童年时听说太阳物质成分可以通过光谱分析来揭示而点燃的。20 岁之前他就曾构思过不止一种可望更有效地分析太阳光的新型光谱仪。这位在达尔文的《物种起源》出版 10 年之后出生的天文学家的终生之梦，正是有一天科学能统一地解释恒星的起源与演化和生命的起源与演化。今天可以说，这个梦想几乎实现了——这正是本书的立足点。在相当大程度上正是由于海尔作为望远镜制造者、资金筹集者和天文台台长的热情和技艺，我们才如此接近于实现他的梦想。

海尔走上建立新一代望远镜和天文台的倡导者之途颇有点偶然，那是在他听说南加利福尼亚大学订购了用于 40 英寸(约 1.02 米)折射望远镜的透镜却又无力付款的时候。望远镜的能力是以其主放大透镜

或反射镜的直径来衡量的。当时，即 19 世纪 90 年代，最大的折射望远镜(即主镜是透镜而非反射镜)是 36 英寸(约 0.91 米)，设置于加州圣何塞附近哈密尔顿山上的利克天文台。这已经接近于制造高精度天文透镜的实际限度，因为更大的透镜会由于自身重量而弯曲变形。当今最大的望远镜全都是反射镜，即用很大的抛物面镜而不是透镜来会聚接收到的恒星光线。反射镜与透镜相比的巨大优越性是，由于没有光透过，镜的背面就可以用架子支撑以保持镜子不变形。在 19 世纪 90 年代人们已开始认识到，望远镜设计的下一步应当考虑大的反射镜，但海尔有兴趣的是得到南加利福尼亚大学订购的透镜并用来建造比利克天文台的那一架更大的望远镜，那些透镜已经造好并放在巴黎那家制造厂的库房里。海尔是富商的儿子，知道如何搞钱，于是赶紧去找芝加哥的其他富家，终于从有轨电车大王叶凯士(Charles Yerkes)那里得到了对所需资金的承诺。那笔钱的总额是349 000美元，是在海尔不懈坚持之下被很不情愿地一点一点吐出来的，前后拖了好几年。望远镜终于造出来了，成了芝加哥大学叶凯士天文台的镇台之宝，海尔于 1897 年被委以首任台长之职，时年 29 岁。40 英寸(约 1.02 米)的叶凯士望远镜现在仍然是世界上最大的折射望远镜。

海尔一发而不可收。虽然设在威斯康星州威廉斯贝的天文台里的这架大型折射望远镜已经非常之好，但他还想要更大更好的，放在更理想的地方。从地上观星的最佳去处是高山顶上，避开低层大气中的尘埃和云雾，远离城市的灯光。海尔来到加州的威尔逊山，住在一间被遗弃的小木屋里，用一架小望远镜来检测天空的清晰度。再度游说之后，他得到华盛顿卡内基研究院的支持，在威尔逊山建一座新天文台，并出任台长，台里最早的设备是一架主镜口径为 60 英寸(1.52 米)的反射望远镜。主镜是他父亲送的礼物，望远镜于 1908 年投入使

用，为确定银河系真实尺度的那个人准备好了主要武器。

沙普利(Harlow Shapley)1885 年出生于密苏里的一个农民家庭，幼年几乎没受正规教育，16 岁时当了堪萨斯一家报纸的刑事新闻记者。但是沙普利认定正规教育会使自己的前程更好，于是在一所迦太基长老会的高等教育学校学习两年之后又想去密苏里大学学新闻学。他于 1907 年到了那里才知道那门课程还要过一年才开设，想想自己为受教育已花费了许多时间，他决定学别的，学什么都行，就是不能闲荡。后来(他 1972 年才去世)他总喜欢说，他选了天文学是因为英文中它的第一个字母是 A，排在开设课程表的最上方，第一眼就看到了。4 年后，他在自己胡乱选择的专业得了学士和硕士学位，然后去了普林斯顿，由罗素(Henry Norris Russell)指派研究双星。此后 3 年里沙普利的主要工作是关于交食双星，他还一劳永逸地确定了造父变星不是双星而是脉动的恒星。1914 年，即发觉密苏里大学没有开设新闻学课程的 7 年之后，沙普利从普林斯顿获得的是博士学位和最优秀的新一代天文学家之一的声誉。他得到了威尔逊山上新天文台里的一个职位，拿到了每月 135 美元的工资，而更重要得多的是，他有了世界上最大的 60 英寸(1.52 米)望远镜。

这是在赫茨普龙首次成功地把造父变星用作距离标尺的一年之后，而对造父变星的研究是沙普利博士论文的一部分，尽管是次要部分。有世界上最好的望远镜在手，沙普利开始用造父变星来测绘银河系。他采取的方式是拣出银河系里的另外一些东西，它们与我们至此已经遇到的都大不相同。那就是球状星团，即一群恒星聚集成球状，每一群里有几万到几百万个成员，即使在中等望远镜的视野里也像是些漂亮的宝石。[4] 球状星团大体上都位于天空的同一个部分，而且看上去又聚成一个球，但那究竟是个离得较近的小球还是个遥远的大球

呢？幸运的是，球状星团里常有造父变星，一个星团里就可以找到好几颗。于是，沙普利就用新标尺和新望远镜测出了一些球状星团的距离。他发现，各个星团里看去最亮的恒星似乎都有着大致一样的真实亮度。这一来他又找到了一种新方法，就连那些没看到造父变星的星团也可以估测出距离了，因为可以认为那些星团里的最亮恒星也和已知距离的星团里的最亮者实际上一样亮，再由视亮度就可以算出其距离。

所有这一切的结果是一幅崭新的银河系图像。众多的球状星团组成一个离我们很远的巨大的球，球心就在银河核心区人马座方向的某一个点上。唯一合理的结论就是，这个球形系统的中心的确也就是银河系的中心，太阳和太阳系是位于这座恒星都市的郊区，即从星城中心到边缘总距离的大约2/3的地方。沙普利的结果是在 1918—1919 年间的几个月中以一系列论文发表的。他把银河系的整体大小高估了将近 3 倍，这是因为他没有考虑尘埃对遥远球状星团的光的遮掩效应（这会使星团看上去变暗，而沙普利也就误以为它更远）。但是他的基本结论是正确的。我们今天知道，银河系正像赖特和康德所想象的那样是个扁平的盘子，直径大约是 100 000 光年（30 000 秒差距）。从盘

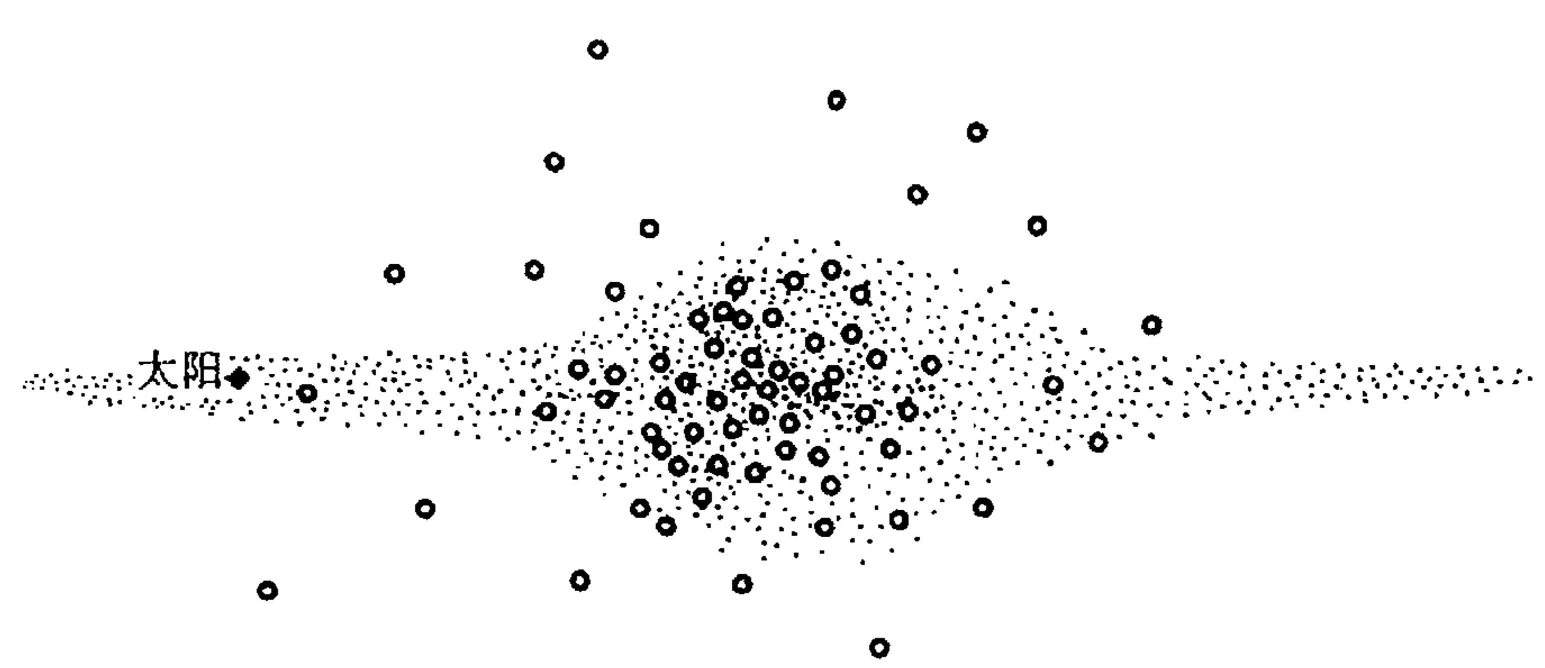


图 2.4　沙普利发现，球状星团（图中以圆圈表示）的分布表明太阳和太阳系是在远离银河系中心的地方。

心到边缘的距离是约50 000光年，太阳距离盘心约30 000光年，所以我们是生活在这个星系的边远地区。但是，我们银河系在宇宙中的重要性如何呢？勒维特、赫茨普龙和沙普利已经把对向“上”能看多远的回答推进到了数十万光年。那么，还能继续向“上”延伸多远呢？

横贯宇宙

在这个路口上，沙普利畏缩了，作了错误的转折。这并不全是他的过失，因为他为整个宇宙构造一幅智力图画、一个想象模型的意图仍有赖于别人所作的观测和对观测的解释。但是这个错误改变了他的整个学术生涯，所以尽管他在同行中成就卓著、颇受尊重，但他后来总是回首1914年至1920年在威尔逊山的日子，那才是他一生事业的巅峰期。

沙普利对银河系尺度的高估使得麦哲伦云看来也只是我们星系的一个组成部分，而不是另外的星系。既然如此，其他星云，例如仙女座那个很大的旋涡状星云，在沙普利看来就必定也在银河系之内，或者至多是银河系的小小的伴星系。按照这幅图景，银河系基本上就是整个宇宙，向“上”已经到了极限。但是还有别的天文学家认为，星云必定也是星系，沙普利一定是把银河系的尺度高估了，尽管他们还不知道为什么会高估。持此观点最力的是利克天文台的柯蒂斯(Heber Curtis)。[5]

柯蒂斯是另外一个沿反常途径走来的天文学家。他1872年生于密歇根州的马斯基根，攻读古典文学并在22岁时当上加州纳帕学院的拉丁语教授。后来他的兴趣转到了天文学上，当纳帕学院与太平洋大学于1897年合并时，他成了天文学和数学教授，这可是今天大学里的任何人听了都会大吃一惊的转折。在不同的天文台做了几次短期

研究之后，柯蒂斯于1902年到了利克并在那里一直干到1920年，其中有一段时间去智利作南天观测。1909年从智利回来后，他全神贯注于确定旋涡星云的本质。拍摄得越来越好的照片终于使柯蒂斯相信，旋涡星云是和我们银河系一样的星系，它们中有些为我们从正面看去，因而整个盘面和旋涡结构都能看到，其他的则是从侧面看或是从某种倾斜的角度看。倘若果真如此，这些银河之外的星系必定距离非常遥远，在望远镜里看去才只是一小块云。但是又怎么测量那些星云的距离呢？对这个问题曾有过两种答案，在1920年前一直难分孰是孰非。答案之一是基于对1885年看到的一次事件的错误理解。不幸的是，沙普利跟着错了。

1885年8月20日，哈特维希(Ernst Hartwig)看到在仙女座星云出现了一颗新的恒星，即新星。这颗星又很快变暗并消失了，幸亏它在最亮时已被抓住。这是第一次看到与该星云有关联的单颗恒星。对这个事件的一种解释是，哈特维希目睹了银河系内一团涡动的气体和尘埃云里一颗新恒星的诞生。不论那颗星是什么，它短暂的光辉有仙女座星云的所有其余部分合在一起那么亮。一个判定究竟有多亮的机会似乎来到了，那是在1901年在英仙座方向上看到另一颗恒星爆发的时候。那颗新星离得很近，用视差法估算出是大约100光年。由于对仙女座星云没有更好的距离估计，那时的天文学家猜想两颗新星应该有同样的真实亮度，于是由视亮度得出前一颗新星的距离是1600光年。这意味着仙女座星云虽然相当大但仍在银河系范围之内。

这基本上就是沙普利的论证，用以支持他关于银河系是宇宙的主体、旋涡星云不过是些小小下属的主张。柯蒂斯相信星云本身都是星系，他也在寻找证据支持自己的观点。假定1885年在仙女座看到的新星实际上比1901年英仙座的那一颗要亮得多，如果仙女座星云像

银河系那样也是一个星系，那颗新星就得在短时间里像 10 亿颗恒星那么亮，这在沙普利看来是荒唐的。谁能说他不对呢？但是我们现在知道，非常罕见的“超新星”的确有那么亮。而使我们得以知道这一点的原因之一，正是柯蒂斯决定寻找仙女座的其他新星并把它们的亮度与 1885 年的那一颗和 1901 年英仙座的那一颗相比较。

柯蒂斯终于在仙女座发现了另外几颗新星(迄今已有 100 多颗记录在案)，这表明仙女座星云必定是许许多多恒星的集合，因为新星并不那么常见。而所有后来的那几颗都比 1885 年那颗要暗得多这一事实又作出启示，应该拿后来几颗同英仙座新星比较。这样重做的结果是使仙女座星云的距离增大了 100 多倍，远在银河系边界之外。究竟谁对，柯蒂斯还是沙普利？问题是如此重要而有趣，所以美国科学院于 1920 年在首都华盛顿组织了一场两人之间的大辩论，爱因斯坦也是听众之一。这场辩论被广泛报道，普遍的感觉是沙普利输了，而柯蒂斯对宇宙尺度的阐述是正确的。[6] 这次失败之后，沙普利立即离开威尔逊山去了哈佛，担任哈佛大学天文台台长，该职位本是 1919 年皮克林去世后提供给他的。尽管沙普利对天文学又作出了许多其他贡献，但他对自己的这一决定必定感到懊悔，因为他看到威尔逊山的一位新人哈勃，正是从他停下来的地方起步。哈勃接过了沙普利的利用造父变星和球状星团估测距离的技术，而且他还得到了一架比 60 英寸(1.52 米)望远镜更大更好的望远镜。

哈勃是海尔的勤奋活动的最大受益者。海尔并不满足于他的威尔逊山天文台里的 60 英寸(1.52 米)望远镜，又说动了一个洛杉矶商人胡克(John D. Hooker)为一架主镜口径 100 英寸(2.54 米)的望远镜掏钱。这架胡克望远镜是 1918 年建成的，是此后 30 年中世界上最大的望远镜。海尔本人心力交瘁，于 1923 年 55 岁时听从医生劝告，辞去

了威尔逊山天文台台长之职。但当他在帕萨迪纳附近的家中过平静的退休生活时，又想着要建造一个小天文台和发明一种新型光谱仪来研究太阳。他又出来筹集资金，要在南半球建一座天文台，但却没有搞成，饱受了一次精神打击。但他很快又振作起来，投入了一项新计划，要建造一架更大的即口径200英寸(5.08米)的望远镜。洛克菲勒基金会于1929年出资600万美元，工程由加州理工学院负责实施，拟将望远镜建于加州的帕洛玛山上，海尔亲任工程规划组的主席。但是，海尔没有亲眼看到自己的这件最大杰作。他于1938年去世，而由于第二次世界大战和其他延误，望远镜的建成用了20年时间。[7]1948年，海尔望远镜投入使用。1969年，威尔逊山和帕洛玛山的两个孪生天文台合名为海尔天文台，以颂扬这位把美国天文学提高到20世纪世界领先地位的伟大人物。而在这次命名的很久以前，哈勃就已经使所有天文学家睁开了双眼，看到了宇宙的真实尺度。

哈勃和沙普利一样也是密苏里州人，1889年生于马什菲尔德城，是当地一位律师的7个子女中的老五。他在芝加哥上了高中和大学，上大学时海尔正在那里当教授。哈勃体格强壮，本有人要他去做一名职业拳击手，去与当时的拳王约翰逊(Jack Johnson)一拼。他却接受了罗兹(Rhodes)奖学金远赴英国牛津大学学法律，成了代表牛津的运动员，还曾作为一名业余拳击手在一次表演赛上与法国冠军卡尔庞捷(Georges Carpentier)较量。他于1913年返回美国并在肯塔基当律师，但是只干了几个月就断定这并非自己喜爱的职业。他在芝加哥大学读书时部分地由海尔激发起来的对天文学的兴趣又复活了，于是他重返母校改读天文学，并在叶凯士天文台做研究助理。他于1917年完成学业并获得博士学位，又由海尔安排了在威尔逊山的一个职位。但他却应征入伍，作为一名步兵去法国打仗，在那里右臂曾被弹片击

伤。1919 年他终于来到威尔逊山，其时正逢新的 100 英寸(2.54 米)望远镜刚投入使用，而沙普利又即将去哈佛。哈勃的机遇真是再好不过了。刚刚在 1917 年才由威尔逊山的里奇(George Ritchey)第一次从照片上证认出一颗新星，这启发了柯蒂斯去重新查看利克天文台以前拍摄的照片，并找到了证据，使河外星云的距离第一次能被直接测量。关于那些星云本质的争论已经延续了数百年，终于在 1924 年结束。100 英寸(2.54 米)望远镜与哈勃的组合，终于为人类绘出了一幅宇宙的新图景，而更惊人的发现正接踵而来。

哈勃的宇宙

哈勃相信那些旋涡星云是远在银河系之外的星系，但他并没有轻率地去作证明。他先从其他星云的问题入手，那些星云并不呈现特征的旋涡结构，几乎肯定是在银河系之内。他用不同的望远镜观测，经常是用 60 英寸(1.52 米)望远镜，起初只能偶尔在允许时间里用 100 英寸(2.54 米)望远镜。他在 1922 年完成的这项研究表明，那些气体星云(还含有尘埃)并不像恒星那样自己发光，而是要么反射它们里面或附近的恒星发出的光，要么就是从附近恒星吸收的能量足以把气体加热到发光。气体星云与银河系内恒星之间的这种联系，证实了那些星云自身的确是银河系的成员。那么，那些旋涡星云是什么呢？哈勃的“训练期”已经结束了，他现在把注意力转到了自己心目中最重要的问题。

即使在 20 世纪 20 年代初，即使有 100 英寸(2.54 米)望远镜，仍然没有迹象表明有可能得到足够好的旋涡星云照片，能清楚地显示星云被分解成一颗颗的恒星，就像麦哲伦云的情况那样。哈勃所能得到的最好照片在放大镜下看去(如果光线适当而且他的情绪乐观的话)，

也只能隐约觉得那淡淡的光似乎可以分裂成颗粒结构，从而暗示所研究的星云是单个恒星的集合。但对谨慎如哈勃者而言，这不是他愿意用以维系自己声名的证据。既然旋涡星云还难以被分解成分离的恒星，哈勃决定先去研究一团能够分解出恒星的云，即使它只是天空中一块暗弱的、不规则的光斑，不像麦哲伦云那样引人注目。他选中了一个名为 NGC(表示“新总表”)6822 的星群，花了两年时间来拍摄最高质量的系列照片。哈勃把那团恒星云锁定在 100 英寸(2.54 米)望远镜的视场里，一小时又一小时地耐心观测，在好的情况下一个夜晚能得到一张有用的照片，在其他情况下则要用两个整晚才能拍出一张还过得去的照片。当然，望远镜还有别的任务，不能全归他来干这一件事。这样，哈勃用了 1923 年和 1924 年两年的绝大部分时间才得到 NGC 6822 的 50 张好照片。他从这些照片上证认出十几颗造父变星，并用沙普利的技术定出这个小小的不规则星系的距离是小麦哲伦云距离的 7 倍，那是在 1924 年。

当 NGC 6822 的观测计划仍在执行中时，另一颗河外造父变星从仙女座星云(又称 M31，即在梅西叶表中列为第 31 号)中被证认出来了。这项发现是在 1923 年秋一次旨在寻找仙女座星云里的新星的巡天观测中作出的，因为新星可以用来检验柯蒂斯关于星云本质的主张。哈勃在他的《星云世界》一书中回忆道：“那个计划中用 100 英寸(2.54 米)反射镜拍摄的第一张好照片，导致了两颗正常新星和一个暗弱天体的发现，后者开始时也被假定为一颗新星。但与威尔逊山的观测者们在以前寻找新星时积累下来的大量系列照片所作的对照，表明那颗暗星是一颗变星，而且很容易指出其变化特征。那是一颗典型的造父变星，周期大约是一个月……所要求的距离是在 900 000 光年的量级。”由于各种原因，对该距离的估测现在已经修改为

2 000 000光年(670 千秒差距)以上。但是与上述发现带来的突破相比，这种修改只是次要细节。完全没有任何新假设(不像柯蒂斯，他只能猜想仙女座星云里的新星与银河系里的新星本质上一样)，只是用沙普利已经用来测量过银河系的同一把尺子，哈勃已能测出比较近的河外星系的距离。

这恐怕是现在值得领会的最重要一点。到 2 000 000 光年的惊人一跃只是进入宇宙的第一步，只是到了像银河系那样的众多其他星系中的最邻近者之一。我们生活于其中的银河系在天文学的图像里突然缩小了，缩成了一粒小尘埃，飘浮在巨大的、黑暗的虚空里。

银河系地位的这种下降还是用了一点时间。开始时我们这个星系似乎显得比别的星系更大更重要。直到 1952 年，由于修正了对造父变星标尺本身的长度，才明白其他星系也像我们的银河系这么大，而且比哈勃估计的还要远。由于有了更好的照相感光乳胶，哈勃在 1923 年成功地把仙女座星云(M31)的外围部分分解成大群密集的恒星，并且后来又从 M31 和另一个距离大致相同的旋涡星云 M33 里证认出更多的造父变星。到 1924 年底，确定旋涡星云本质的证据已经很充足，并由哈勃提交美国天文学会的一次会议。在此后 5 年里哈勃又积累了更多的证据，1929 年他对这些证据作了明确的总结。同时，他还开始发展更新的技术，用以估测那些远在单颗恒星能被辨认、造父变星标尺能够使用的范围之外的星云(即星系)的距离。

即使用 200 英寸(5.08 米)望远镜，也只能在大约 30 个最近的星系里证认出造父变星。如果一切顺利，计划中做轨道运行的空间望远镜*必将对此作出改进，但是测量更遥远星系距离的其他技术终究是

---

* 指现已投入使用的哈勃空间望远镜。——译者

需要的。哈勃的第一个步骤又是学自沙普利，就是用超巨星来做距离指示器，正如曾用此种恒星来指示银河系中球状星团的距离那样。哈勃用此法能测到的距离为所能看到的造父变星距离的 4 倍，即大约 1000 万光年。球状星团可以用作而且已被用作宇宙的一把粗略量尺，其前提是假定每个星系里最亮的星团实际上也同银河系里最亮的星团一样亮，但现在天文学家可是开始倾箱倒箧来寻找测量越来越远的星系距离的办法了。为着向前推进，哈勃不得不作了一个大胆的、只有粗略精度的假定。当他注目室女座方向的一个大星系团时，他发现那些星系的亮度互相都差不多，最亮者的亮度也只是最暗者的 10 倍。他假定所有那些星系都有同样的真实亮度，即为最暗的那个的 3 倍，或最亮的那个的 1/3，从而估计出了距离，并且有理由相信所得结果与正确答案的差异不超过 3 倍——也许是 3 倍，也许是 1/3，但不会更糟。这种技术后来又通过只选用一个星系团中的最亮星系为标准而得到改进。(因为最亮的星系的确相互都差不多，最亮的超巨星也是如此。)此法尽管是近似的，却把哈勃送到了大约 5 亿光年的距离。[8] 在这样一个空间范围里有大约 1 亿个星系。但是，所有这些距离测量都依赖于造父变星标尺的最初定标，这种定标是在银河系以内，更准确地说是在太阳附近，用统计视差技术(和现在的颜色—星等图)作出的。我们关于宇宙距离的丰富信息就像一座倒立的金字塔，金字塔在向上和向外膨胀，而它赖以平衡的基点就是已定标的造父变星标尺。不测出那 30 多个已在其中证认出造父变星的星系的距离，就根本无法定标其他虽然更粗糙但尚可使用的标尺。如果造父变星标尺定标错了，但后来又修正了，我们就不得不改变宇宙的整个尺度。如我们将要看到的，这在过去几十年里已经发生了好几次，最重要的一次是在 20 世纪 50 年代初。但是，所有这些修正都没有改变哈

勃所建立的宇宙基本图景。

哈勃的宇宙，也就是我们的宇宙，延伸到几亿或几十亿光年。我们今天用 100 英寸(2.54 米)或 200 英寸(5.08 米)望远镜看到的星系中，有些是如此遥远，以至于我们接收到的光竟是在地球形成之前发出的。人的大脑的确无法领会宇宙的尺度。我们所能做的一切就是盯着数字。数字告诉我们，即使是 M33 和 M31 这两位近邻，光都要用 200 万年或更长的时间才能越过它们与银河系之间的天堑，我们听了的确感到茫然。即使是最伟大的宇宙学家，一个爱因斯坦或者一个霍金，其内心深处都必定对卡莱尔(Carlyle)的话颇有同感："我决不妄求了解宇宙，它可比我大得太多了。"

但是，天文学家仍然试图了解宇宙，尽他们所能去了解。哈勃奠定了现代宇宙学的基础。他不仅确定了宇宙的尺度，而且描述和划分了星系的主要类型——已知的星系中 75%是旋涡星系，其余的差不多都是雪茄状，或者说橄榄球状的椭圆星系，只有很少数是不规则星系(最后这种类型也许是因为太小太暗而看不清楚)。此外，他还分析了星系的空间分布，发现总的说来分布是均匀的。虽然星系聚集成团，星系团在空中却是随机分布的，在天空中一个区域看到一个星系或星系团的机会与在另一个区域看到的机会一样多，当然必须把银河系里尘埃的遮掩效应考虑进去。这的确是一个重大发现，因为它意味着宇宙的终极结构样式可能已被探明。这一发现的重要性最近被稍稍减小，因为有证据显示可能还要多一个结构层次，即星系团组成的团。仍然具有根本意义的是，宇宙在所有各个方向上都是同样的，宇宙中没有任何特殊的区域。但即使是如此重要的观测结果与 1929 年哈勃那项出人意料的发现相比也黯然失色。他发现，所有这千百万个星系都在相互分离，都在以最高可达几分之一光速的巨大速度互相分开。

宇宙整个儿地在膨胀，这就清楚地指出，整个宇宙必定有某个确定的时间起点。天文学家往黑暗的夜空里能看多远似乎并没有什么限度，但是宇宙的膨胀却意味着，对宇宙的历史能回溯多久却有一个限度。正是宇宙膨胀这一近至 1929 年才作出的发现，才真正把天文学家送上了通往大爆炸之路。

注　释：

1. 更准确的值是 31 角分。

2. 如果一个人看到月亮在头顶上，另一个人在同一时刻看到它在地平线上，这两人视线之间的视差移动就是 57 角分，差不多是月亮角直径的两倍，而相应的三角形的基线就等于地球的半径，于是就可以得出月亮的距离。

3. 按波长减小的次序排列的颜色是红、橙、黄、绿、蓝、靛、紫。

4. 这种星团里的恒星在望远镜里看去紧紧挤在一起，与星团以外的恒星相比它们也的确相互很靠近，但是它们之间仍有相当大的空间。在这种星团密集的核心区，恒星之间的平均距离大约是太阳附近恒星之间距离的 1/10。

5. 并不出人意料的是，加利福尼亚州的这两个大天文台之间曾在许多年里有过激烈的、有时带点痛苦的竞争，利克天文台属于加州大学，而威尔逊山天文台属于卡内基研究院，后来属于加州理工学院。当然，现在双方早已和解。

6. 颇有点令人啼笑皆非的是，虽然柯蒂斯对仙女座星系距离的认识是对的，但他对银河系的观点却几乎全错了，他认为太阳是在银河系的中心，他也大大低估了银河系的尺度。1920 年所能得到的关于宇宙的最好图像应该是把沙普利的一半和柯蒂斯的一半拼起来。

7. 这项工程的副产品也对美国乃至全世界的天文学起了重要作用：负责制作 200 英寸(5.08 米)反射镜的纽约科宁玻璃公司，决定先做一系列较小的镜子来逐步提高技术，那些小镜子都没有浪费，两块 61 英寸(1.55 米)的到了哈佛的沙普利那里，一块 76 英寸(1.93 米)的到了多伦多，一块 82 英寸(2.08 米)的到了得克萨斯州的麦克唐纳天文台，一块 98 英寸(2.45 米)的到了密歇根大学天文台，还有一块 120 英寸(3.05 米)的到了利克天文台。

8. 这个值是哈勃自己认为的，现在已被修正得大了许多，但你仍可从中品出他快马加鞭的滋味。

第三章　膨胀着的宇宙

科学并不总是以某种有序的方式进展。今天的一个发现，可能要等几年或几十年才能明白其意义，并给予它应有的地位，而明天得出的另一个结果，却可能立即显示出重要性。一代或几代人会沿着不同的研究路线仿佛各自独立地前进，直到某种关联因素表明他们看到的是一个更大整体的不同方面。但是，与哈勃密切相关的两条研究主线却并非如此。从星系的巨大速度一被发现起就很明显，这将对宇宙的本质作出重要的揭示，哈勃对暗弱而遥远的星系之数目与分布的普查亦是这样。但这两个前沿上的进展是间歇的、交替的，关于研究历程的记录就不得不先沿一条轨迹走一段，然后再返回到另一个主题。

宇宙中均匀地分布着星系团，在所有方向上用望远镜尽其所能看去都是一样的，证明宇宙结构的这种基本性质的工作在 20 世纪 30 年代及以后都在继续，并且一直延续到今天。哈勃的最初期望之一，如他的同事已讲过多次的，是用 100 英寸(2.54 米)望远镜拍摄到夜空某个部分的照片，其中能看到的星系将如作为前景的银河系中的恒星那

样多。这个期望于 1934 年 3 月 8 日实现了，证实宇宙中的星系确实如同银河系中的恒星那样多。那一天是一个里程碑，宇宙的广度和星系作为其基本可见单元的本质这两点，终于都排除了并非无理的怀疑而得到确认。哈勃通过对这类照片上星系的计数得出，原则上应有 1 亿个星系能被拍摄到。天文学家更进一步计算出，如果愿意并有时间对整个天空作详细普查的话，则由 200 英寸(5.08 米)的和其他现有的大望远镜能找到 10 亿个星系，这当然如银河系中的恒星那样是大约而言。但是在 1934 年得到那张里程碑式照片之前 5 年，哈勃已经报告了自己的另一项发现：除了最邻近的几个以外，所有星系都不仅随宇宙的膨胀而离我们远去，而且遵循着某种简单的物理定律。而在这一发现之前 17 年，即 1912 年，才首次对当时仍称为“星云”的东西测量了速度。

红移与蓝移

关于遥远星系红移的故事，实际上始于 19 世纪波士顿的一位富家子弟洛厄尔(Percival Lowell)对那颗红色的行星——火星的迷恋。洛厄尔生于 1855 年，在哈佛大学学数学并于 1876 年毕业。一年后，意大利天文学家斯基帕雷利(Giovanni Schiaparelli)报告了他对火星上“卡纳里”(canali)的首批详细观测。“卡纳里”的意思是“水道”，斯基帕雷利借用这个词，以及借用“海”、“大陆”等词，纯粹是用于描述火星上的特征，而不是指真有像地球上这样的海、大陆和河道(更不必说运河了)。但是，部分是由于把“卡纳里”错译成 canals(运河)，部分是出于想当然，斯基帕雷利的报告在法国、英国和北美激起了长达数十年的热潮。许多认真的天文学家，更不用说大量的普通百姓了，都相信火星上有智慧生命，火星人正忙着开运河把水从两极引

到赤道。对斯基帕雷利报告的这种误导之顶点，我想是奥森·韦尔斯(Orson Welles)根据威尔斯(H. G. Wells)的科幻小说《大战火星人》改编的著名广播节目，那小说是在19世纪90年代的火星热时写的。那个广播节目在1938年播出，以真实新闻报道的形式来描述火星侵略者对美国新泽西州的进攻，一时间竟使数以千计不知道那是幻想小说的听众大为恐慌。

但所有这些都还是半个多世纪以后的事，斯基帕雷利的发现以断章取义和误译的形式传到美国并引起青年洛厄尔的注意是在19世纪70年代后期。种子播下后要很长时间才会结出果实。洛厄尔毕业后花了1年去旅游，花了6年参与他父亲的棉花生意，然后是差不多10年在日本和远东。直到1893年返回美国时他才决定要认真搞天文，特别是研究行星。他自有办法在亚利桑那州的弗拉格斯塔夫建一座自己的天文台，那地方空气清净、海拔2000米以上，又远离任何大城市。他用一架24英寸(61厘米)折射望远镜研究火星达15年之久，向热切期望的人们报告了不仅是运河还有绿洲和清楚的植被迹象。这些“发现”自应归功于他的想象力，但是其他天文学家也犯过类似的错误，要知道即使是地球上最好的望远镜也只能给出很差的火星图像，这是因为图像的细节总是因地球大气而变得模糊，同时大望远镜又总是放大了这种模糊作用。洛厄尔虽然在火星的生命问题上搞错了，却确实激发了一代美国人对天文学的兴趣。他还预言在海王星轨道之外必定有第九颗行星，因为其他外行星的轨道似乎受到这个未知行星的摄动。冥王星确实就在他预言的位置上被找到了，但那是在1930年，即他死后14年。这也许是一次侥幸成功，因为现在有的天文学家认为冥王星太小，不足以对外行星轨道产生可观测到的影响，也许还有第十颗行星真正应该对这种摄动负责。尽管如此，洛厄尔的成果

确实而丰富，为学术界公认已远远超过一位富有的业余天文爱好者所能做的。的确，他在 1902 年被麻省理工学院聘为非常驻的天文学教授，讲授系列课程，并保有这个职位直到去世。但是，洛厄尔对天文学的最大贡献或许是雇用了一位名叫斯里弗（Vesto Slipher）的观测者，派他去拍摄旋涡星云的光谱并寻找谱线的多普勒频移。洛厄尔的动机仍是对行星的兴趣，他像当时的许多天文学家一样，认为星云可能是正在形成中的行星系统。但重要的不是动机，而是结果。

斯里弗在许多方面正与洛厄尔相对立。洛厄尔火热、外向，急于得出结论，斯里弗则安静、有条理、刻苦，在完全搞清楚之前绝不宣布自己的发现。他俩性格上的差异是如此显著，以至于有人猜想，洛厄尔正是知道自己的长处和短处，才有意挑选斯里弗到身边来，以便自己冲动时有人浇浇凉水。

斯里弗 1875 年生于印第安纳州的马尔伯里，1901 年毕业于印第安纳大学，并立即由洛厄尔本人邀请进入洛厄尔天文台。他此后的职业生涯就在那里度过，其间于 1903 年和 1909 年先后获得硕士和博士学位，都由印第安纳大学授予，在 1916 年洛厄尔去世后任天文台代理台长，在 1926 年成为台长。正是由他开始的寻找导致了 1930 年发现冥王星。他虽然在 1952 年退休，却一直活到 1969 年，一生经历从天文学家还认为银河系即整个宇宙的时代起，一直到射电星系、类星体和被认为是大爆炸遗迹的微波背景辐射的发现。那个预言背景辐射存在的理论，其由来可以追溯到 1912 年斯里弗对仙女座星云光谱的多普勒频移的测量，我们现在知道那个星云是一个最邻近银河系的大星系。

请记住，多普勒频移是运动光源的光谱里明线或暗线的移动。如果光源朝向我们运动，移动就朝向可见光谱的蓝色一端，称为蓝移。

反之，若光源离开我们而去，则为红移。与静止光源光谱里对应谱线的位置作比较而得到的移动量，能给出对光源朝向或背离我们运动的速度(即多普勒速度)的直接量度。

值得谈一下首次测量星云多普勒速度时所用的技术手段。24 英寸(61 厘米)望远镜是一台很好的设备，是当时世界上最好的之一，那时由海尔发动的望远镜技术的大飞跃尚未到来。采用当时最好的技术，即最好的分光镜和最好的照相底片，斯里弗仍需将底片曝光 20 小时、30 小时、甚至 40 小时(当然也就是几个夜晚)才能得到一幅可用于测量多普勒频移的光谱。要知道这一切都是在寒冷的高山上做的，望远镜圆顶里不能用什么取暖器，因为热空气会产生对流，从而使望远镜里和摄谱仪狭缝里的图像模糊；还要总闭着一只眼，用一只眼盯着望远镜里的像，保证那个正被拍摄的星云精确地停留在视场中心位置。当斯里弗得到了照片时，他要做的事还多着呢。来自恒星的光集中在望远镜所成的像的那一点上，即使在这个光点被分光镜展开时，铺开的像仍然亮得足以使谱线能被证认，使谱线的位移(如果有的话)能不太费力地测量出来。但是，星云(指的是星系)的像本就暗弱，经过分光镜展开后就更是如此，很难分辨和证认其中的谱线。如果像展开得太大，谱线就会太弱而根本看不出来；如果要使谱线亮到足以显露，像就不能铺得太开，谱线的移动也就根本测不出来。多普勒测量技术成功的关键是用以记录天体像的感光乳剂的效率，或者说速度。[1] 尽管困难重重，斯里弗还是在 1912 年得到了仙女座星云 M31 的 4 张光谱图，它们全都清楚地显示出对应着每秒 300 千米速度的多普勒蓝移。这就是说，那个星云不仅是在趋近我们，而且其速度大于当时已知的任何天体——恒星、行星或别的什么——的运动速度。

一旦取得突破，其他几个星云的多普勒速度很快也就测出来了，

虽说并非那么轻而易举。斯里弗继续埋头苦干，把他的设备的作用发挥到最大限度，到 1914 年测出了多普勒频移的星云已达 13 个。这时，一幅图像开始浮现。这 13 个星云中只有两个表现出蓝移，其他 11 个都是红移，表明那些星云都在以每秒数百千米的速度远离我们而去。当然，斯里弗测量的首批光谱中红移占优势的情况仍有可能只是一种巧合。但是，你在反复掷一枚硬币时本来预期它的两面出现的次数应该大致一样，结果掷了 13 次有一面只出现两次毕竟会使你惊讶。重要的是，随着越来越多的多普勒频移被测出，红移的优势也越来越显著。斯里弗测出了 41 个星云的多普勒频移，其他天文学家又给增加了 4 个(这也显示了斯里弗的成就，他一个人测的是其他所有人加在一起的 10 倍)，这 45 个中有 43 个显示红移，退行速度最高达每秒 1000 千米以上。这看来就远不是巧合了，尽管那时的天文学家还没有得到最终的证据来表明星云本身就是远在银河系之外的星系。但他们还是有猜疑的。英国的大天文学家也是科普大家爱丁顿，于 1923 年写道：

> 宇宙学中最令人困惑的问题之一是旋涡星云的巨大速度。它们的视向速度平均约为每秒 600 千米，而且退离太阳系的运动占明显优势。一般都认为那些星云是已知最遥远的天体(尽管这个观点被某些权威反对)，所以这里就是我们有可能从中寻找世界的总体性质的地方，如果真有这种地方的话。

这里的“世界”显然是指我们现在所称的宇宙。他继续写道：

> 正向(即退行)速度的巨大优势是非常引人注目的；但不幸的

是还缺少对南天星云的观测，因而还不能得出最后结论。[2]

在此后两三年里对多普勒测量作出了一个重要的修正，那是由于银河系被证实在作整体转动，而且可以用各种新技术估测太阳围绕银河系中心旋转的速度。结果表明，太阳系以大约每秒250千米的速度差不多恰好朝正对着仙女座星云的方向运动。那给人深刻印象的每秒300千米的蓝移实际上主要源于我们自己绕着银河系中心的运动，只有大约每秒50千米才真正是仙女座星云朝向银河系的运动。而测到的红移却表明了更高的退行速度，那两个蓝移也就退到了比最初印象更低得多的地位。舞台已经为哈勃和他的同事赫马森（Milton Humason）造好了，他们将为我们展示宇宙的第一幅现代图景。

红移主宰

20世纪20年代中期，有些天文学家，其中最有名的是德国人维尔茨(Carl Wirtz)，怀疑斯里弗测量到的最大退行速度是否属于所研究的最遥远的星云。但也只能是怀疑，因为在哈勃之前没有人真知道星云的距离。于是自然就该是哈勃来把红移和距离这两方面的证据放在一起，并从而得出红移—距离关系。在确定太阳绕银心旋转的速度以后，部分地由于维尔茨的工作，1927年时在已测出红移的40多个星系中已能看出这种关系，而这又推动了一个新的重大计划来测量更暗弱更遥远的星系的红移，这个计划主要是由赫马森来执行的。

维尔茨在1924年已经得出，从地球上看去的星云视直径似乎与它们的退行速度相关联。他有42个星系的资料，并发现看上去越小的星系其红移很可能越大。假定所有星系的实际大小都差不多一样，这立即就使人想到，较小的星云只是看上去小，因为它们离得更远，

于是大距离与高退行速度就连在一起了。但这只是一种约略的估量，因为那时还没有直接测定星云的绝对距离。这种状况一直持续到1929年。那时斯里弗的注意力已经转到了别的问题上，已经知道红移的星系只有46个。但是那时候已经清楚它们的确是星系，而且哈勃对勒维特和沙普利的开创性工作的发展已经使自己对星系的相对距离有了很好的认识。尽管我们现在知道他的标尺有误，他仍能完全准确地说出一个星系与我们的距离是另一个星系的2倍或1.5倍，或者其他任何比率，而这也就是他所需要的全部信息。即使如此，哈勃从很少的资料中提取出能延伸很远的正确结论仍是令人惊叹的。

哈勃手中有46个星系的红移，基本上得自斯里弗，除此之外他只有18个星系和室女座星系团的距离。比较这19个天体的红移和距离的一个显而易见的办法是作一张速度(红移)对距离的图。每个天体都有一个速度和一个距离，所以就对应着图上的一个点。当所有这19个点都标出来时，哈勃断定它们是处在一条直线上，这就意味着速度必定与距离成正比——假如一个星系的距离是另一个的两倍，则前者的退行速度也是后者的两倍。这其实是一个推广到大范围才正确的结果，而作为其基础的由哈勃标到图上的那少量点子是相当弥散的(见图3.1A)。画一条直线穿过那些点子并且说那些点子是落在这条线上，所需要的是信念而不是想象。但是天文学家习惯于作此类冒险游戏，而且如我们将要看到的，哈勃对他所寻找的关系很可能事先已经有了想法。尽管他得出这个关系的依据在今天看来真是摇摇晃晃，后继的研究却很快就排除了所有的怀疑而证明它在本质上是正确的，今天它被称为哈勃定律：红移与距离成正比。如果这条定律确实是普适的，并假定比例常数已被正确定出，它就给了天文学家测量宇宙的最终标尺。他们所须做的一切只是测量红移，然后就能知道距离。

这个发现的意义还远不止于此，但在深谈那些意义之前应该先赞扬一个人，是他在斯里弗离开的地方继续了红移研究，也是他作出了甚至比哈勃本人更多的贡献来确立“哈勃定律”的有效性。

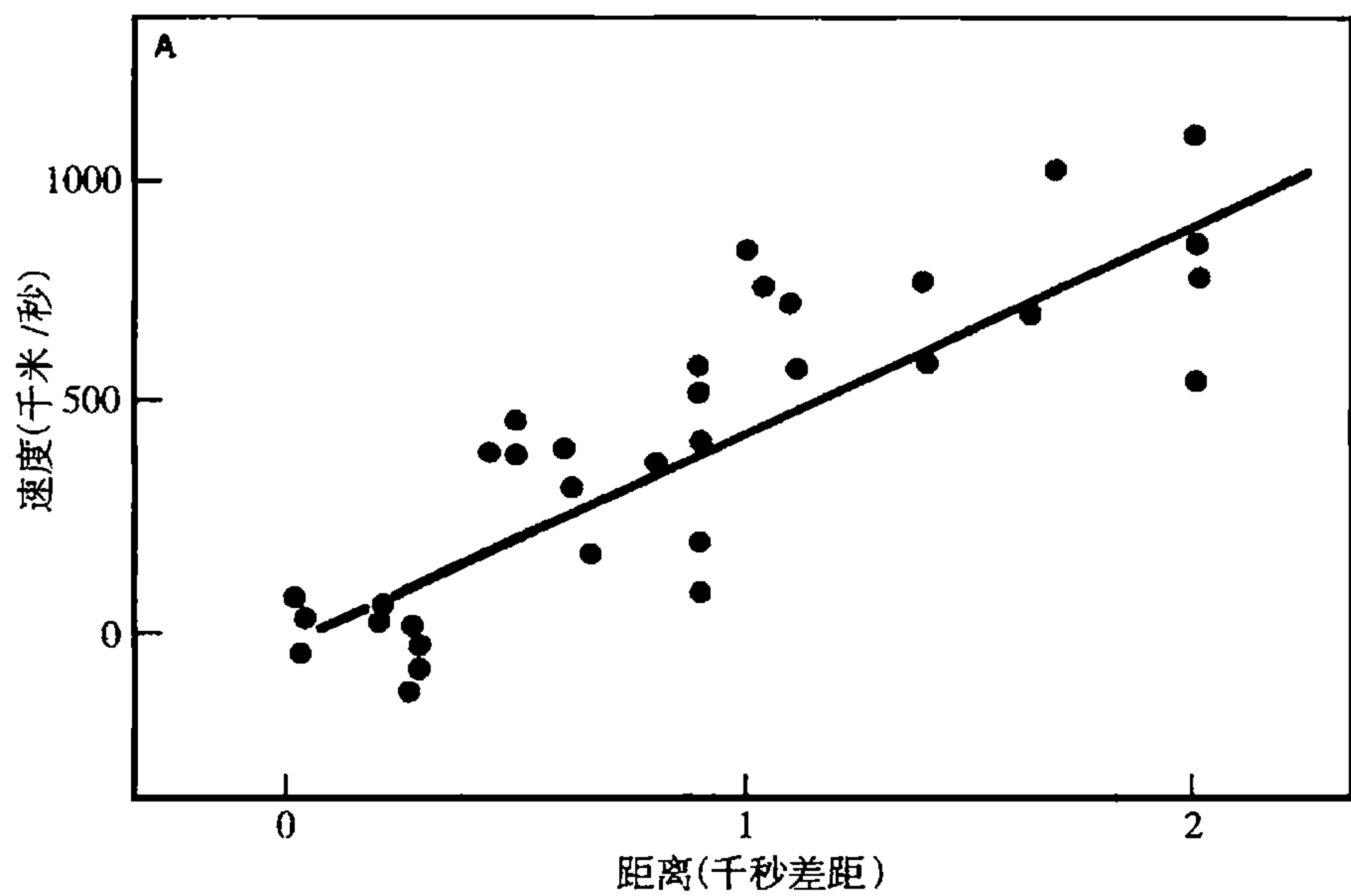


图 3.1A 哈勃最早的红移—距离图之一，完成于 1929 年，其中只有 33 个资料点和贯穿这些点子的一条相当乐观的直线。

赫马森 1891 年生于明尼苏达州的道奇森特。14 岁时参加了一次威尔逊山上的夏令营，他太喜爱那座山了，回家后几天就说服父母同意，离校一年重返山上。他再也没有回去接受正规教育，而是沿着一条迂回途径成为他那一代人中一流的观测天文学家。这个退学学生有一段时间当了骡夫，在修建威尔逊山天文台时往山上拉器材。他对那座山和天文台的工作都很着迷，不过还是有时间与天文台一位工程师的女儿恋爱并于 1911 年结婚。也许是意识到已婚男人的新责任，他决定不再赶骡子，而是到拉文去经营他亲戚的一个牧场。但是在 1917 年，当威尔逊山天文台一个看门人职位空缺时，岳父劝他去干并告诉他，一个爱山又爱天文台的好小伙子应该把这项工作作为进身之阶。不过这位 26 岁的看门人恐怕不会想到，他将在那个台阶上攀登

数十年。

看门人赫马森很快就被提升为夜间助理，任务是照料望远镜并给作观测的天文学家打杂。任何称职的夜间助理都会很快就自己试试做些观测，赫马森则显示出了使用望远镜的才干，于是在1919年被聘为助理研究员，成了天文台研究队伍中的年轻一员。海尔不得不为此项聘任而力排众议，毕竟赫马森是一名骡夫和看门人，14岁以后就没有受正规教育，甚至他同那位工程师的女儿结婚都被一些人看作是谋求提升的不正当手段。但是海尔知人善任，坚持己见。赫马森从1947年起担任天文台秘书，负责公共关系和各种行政事务，1954年成为威尔逊山和帕洛玛天文台的研究员。他得到了一些名誉学位，活到1972年81岁生日前的几个星期。他凭借对精密仪器的细心操纵和对大望远镜的熟练技能为宇宙学家提供了基础资料，使他们得以首次构造出详细而富有想象力的宇宙模型，并且追溯到大爆炸本身。这一切都开始于1928年，即哈勃首次指导赫马森测量暗弱的遥远星系的红移之时。

这种观测需要新的仪器和新的摄影技术，更需要几乎为赫马森所独有的耐心与技巧的结合，他在好几个夜晚一连几小时守在望远镜旁，精确地操纵着，为的是能在只有半英寸(约1.3厘米)宽的底片上得到一个遥远星系的光谱。到1935年他已给斯里弗的清单增添了150个红移，其中的最高纪录对应着每秒40 000千米以上的退行速度，这已高于光速的1/8。有了200英寸(5.08米)望远镜后，他继续向更深的空间前进，到20世纪50年代末已把速度纪录提到每秒100 000千米即光速的1/3以上，对应的距离达几十亿光年。自20世纪20年代末以来已经越来越清楚，宇宙的确很大，星系或星系团只是它的构成单元，而且红移在主宰。宇宙在发狂似地向外膨胀。

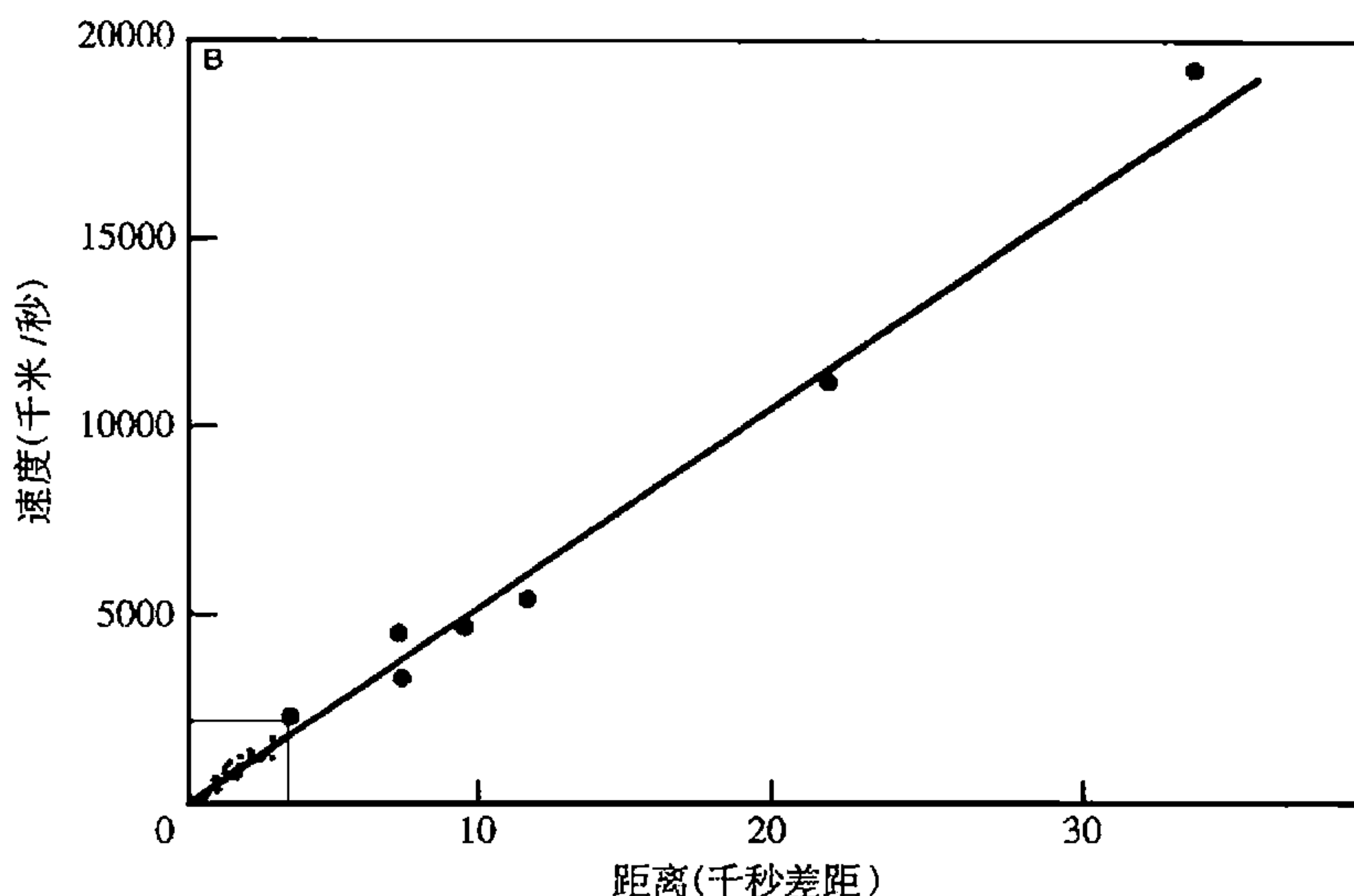


图 3.1B 哈勃的猜测与科学的混合产物的正确性到 1931 年已被证实。哈勃和赫马森大大推进了他们的红移研究，图 3.1A 中的资料现在全部缩进本图左下角的方框里，两张图完成的时间仅隔两年。那条直线仍在，而且看上去要可信得多。

这一切究竟是什么意思呢？该轮到理论家到聚光灯下待一会儿了：他们多少有点忸怩地承认，他们本可预先告诉你宇宙是这个样子，他们只是在近 10 年里才对自己的理论有了足够的信心。这是一个在我们循踪大爆炸的探索时将会熟知的经历。与通常所相信的相反，理论家，当然是宇宙学理论家，似乎对自己的主意没有多少信心。他们并不大声疾呼，而只是缩在一些专业刊物里，因而常常多年不受注意。[3] 从 20 世纪 30 年代到 80 年代，使理论家一再感到意外的是，那些观测新发现与某些人曾在 10 年或 20 年前半心半意地预言过或曾经想到又忽视了的东西竟然完全一致。谁是头号半心半意的理论宇宙学家？谁不相信自己的理论告诉自己的东西？此人不是别人，正是爱因斯坦。

爱因斯坦

爱因斯坦几乎已成为一尊神像，成为当代民间传说中的角色。他

是天才的原型，是个一头白发、有点古怪但和蔼可亲的老人，善于用儿童般天真的话语和提些明显得没有别人会想到该提的问题来直指复杂难题的核心。这些在很大程度上是真实的，同样真实的是他当学生时并不显得多么聪明，没有给学术界留下什么特别印象，因而不得不在20世纪初在一间专利机构当技术鉴定员，但他竟用业余时间发展出了物理学中三个重要的新思想。不过，对这位为科学作出了这些革命性贡献的人物的陈规印象有一方面并不真实。在20世纪头几年，爱因斯坦并不是一个穿着只求舒适不讲仪表、有时嫌麻烦连袜子都不穿，而又可敬可亲的白发老人。从那时的照片看，他是一个黑发的英俊青年，穿着潇洒入时。这一点很重要，因为爱因斯坦最伟大的思想是富于青春活力的思想，这些思想作出了新的洞察，推翻了已建立的观念，真正是富于革命性的。他那使自己受到科学界广泛注意的活力是在1905年迸发的，那时他只有26岁；他最伟大的成果，即广义相对论，则发表于此后10年。虽然他成了民间皆知的富有天才而和蔼可亲的老教授并活到1955年，他最重要的工作却都是在第一次世界大战结束前完成的。科学，尤其是数理科学，就是这样。只有年轻的头脑才能放得开，去发现和抓住新概念，如果新概念是像爱因斯坦提出的那些一样与旧概念截然不同，那就需要你用自己的全部余生，甚至需要几代人，去挖掘其含义。

爱因斯坦作为一个并不突出的学生于1900年毕业于瑞士苏黎世联邦工业大学。他的学习经历已是颇多波折——他到3岁才会讲话，虽然各科成绩有好有差，在15岁时还是以“破坏性影响”的罪名从高中(慕尼黑大学预科学校)被开除。那次开除也许是他故意造成的，因为他父母已因生意失败而离开德国移居意大利，而少年爱因斯坦对德国社会的军国主义性质深感厌恶，大约就在那时放弃了德国国籍，

宁可做一个无国籍者。和家人一起在米兰过了一年自由快乐的日子后，他于1895年报考苏黎世联邦工业大学，但没有考取。经过在瑞士阿劳中学的一年填鸭式学习后，他第二次报考顺利过关，于1896年秋即17岁半时人读苏黎世联邦工业大学。

就在阿劳中学的那一年，爱因斯坦开始苦思一个将在10年后把他引到狭义相对论的问题。他想的是，如果你跑得快到能赶上光，那么在你看来光波是什么样子呢？这的确就是对宇宙作孩提般天真探索的一个最好例子，这种探索方式正是他的个人标志。这是一个可笑的问题，难道不是吗？就像一个三岁孩子在问“为什么草是绿的呢？”但是且慢，也许这个问题比乍听之下要多点什么。

更多的是因为厌倦而不是没有能力，爱因斯坦在大学的考试仅仅勉强及格——他不耐烦去学些自己没兴趣的东西。他这个学生的傲慢使老师们不舒服，据说其中一位，韦伯(Heinrich Weber)，曾对他大叫：“你是个聪明的家伙，可你有个毛病，你不让别人告诉你一点什么，你不让别人告诉你一点什么！”[4] 由于本可帮助他的人与他关系疏远，也由于他学习成绩平平，爱因斯坦毕业后找不到一个学术职位，只好做家庭教师，直到1902年才得到那份众所周知的在伯尔尼专利局的工作并加入瑞士国籍。那份工作很稳定，他干起来很容易，他有了安全感也有了大量时间去思考诸如光波本质之类的难题。但是绝没有人预期到，他那些天才思想的迸发竟是在此后仅仅三年之内。

爱因斯坦的理论

德国杂志《物理学杂志》第17卷于1905年出版，它曾使众多学者惊讶，现在成了收藏家的珍品。在那卷杂志里，年轻的爱因斯坦，一个默默无闻的专利局职员(他那时甚至还没有博士学位)，发表了三

篇论文，给出了对世界本质的至关重要的新见解。一篇有助于确认原子的真实存在，第二篇则提示光可能并不简单地只是一种波，而是还表现得像一系列的粒子。这两篇文章都已被证明对量子物理学的发展非常重要，而爱因斯坦也因第二篇即关于光电效应的论文而于 17 年后获得诺贝尔奖。[5] 但是最著名的是第三篇。它有 30 页，用的是一个并不吸引人的题目“论运动物体的电动力学”。正是这篇文章告诉我们，时间和空间都不是绝对的，它们可以随观测者状态的不同而被压缩或伸张；运动的物体会变重；$E=mc^2$；还有通向原子弹和核电站以及认识是什么使太阳和其他恒星内部保持高温的道路。如果这些听起来已给人深刻印象，请记住爱因斯坦最伟大的工作还在此 10 年之后。

作为狭义相对论基石的 1905 年论文，在今天的世界里倒是足够重要，但在通向大爆炸的道路上只是一个小插曲。爱因斯坦对光本质的思索源于 19 世纪伟大的苏格兰物理学家麦克斯韦(James Clerk Maxwell)的工作，后者建立了把光表述为以一定速度运动的电磁波的方程组，该速度通常被记为 c。关于光波在你以速度 c 和它并肩前进时看去是什么样子的问题，实际上揭示了光的行为与我们从日常经验中得来的“常识”规则之间的矛盾。如果你跑得像爱因斯坦想象的那么快，你将会看到电磁波仍然在运动，但是按照“常识”它似乎不会再运动，这就与麦克斯韦方程组相矛盾。我们对这个世界的认识必定有什么地方错了。麦克斯韦方程组与以日常经验为基础的先入之见是不一致的，这个是非必须搞清楚。爱因斯坦的天才就在于接受麦克斯韦方程组而抛弃那些偏见，从而给出了对真实世界的更好的新表述。

到了 20 世纪头几年，所有测量光速的实验总是都给出同样的结果 c。科学史家仍然在争论爱因斯坦本人当时是否知道这些实验，但这并不重要。通过巧妙地安排光束和镜子，可以测量出与地球运动相

同方向和相反方向上光的运动速度。常识会使你觉得这两个速度应该不同。如果一辆巴士在以每小时 10 英里(16 千米)的速度开走，而我以每小时 9 英里(14.5 千米)的速度在后面徒劳地追赶，巴士相对于我的速度就只是每小时 1 英里(1.6 千米)；如果我坐在一辆速度为每小时 30 英里(48.3 千米)的巴士里，而在高速公路的另一侧有一辆巴士也以每小时 30 英里(48.3 千米)的速度逆向开来，则对我来说后一辆巴士是在以每小时 60 英里(96.6 千米)的速度运动。但是光却不是这样。地球以一定的速度在空中运动，我们把这个速度记为 v。以速度 c 越过我们的光束被我们测到的速度并不是 $c-v$，从反方向迎着我们而来的光束的速度也不是 $c+v$。不论我们自己的速度多大，也不论光束是来自何方向，当我们测量其速度时所得的结果总是 c。[6]

所以，爱因斯坦说，我们必须抛弃日常偏见。对速度而言，1 加 1 并不一定等于 2。他设计了一套数学框架，使得任何在以恒定速度沿直线运动的参考系里的观测者所测量的光速总是一样。这些参考系相互之间可以有相对运动(这就是“相对论”中“相对”一词的由来)，但是不能做转动或加速运动(所以是“狭义”，就是说该理论只能处理一定的物理问题)。在任何一个这种参考系里的人都会得出同样的物理定律，都有权把自己所在的参考系看作是“静止”的，他们所测到的光速都是 c。宇宙中没有特殊的参考系。[7]

这里不再细述，只把爱因斯坦的计算结果简单归纳于下。将两个速度 v_1 和 v_2 相加的改进的规则不是给出 $V=(v_1+v_2)$，而是 V 等于(v_1+v_2)再除以$(1+v_1v_2/c^2)$，这里 c 是光速。由于 c 很大，每秒 300 000 千米，对日常速度如每小时 10 英里(16.1 千米)或 30 英里(48.3 千米)而言，上述除数几乎就等于 1，因为 v_1v_2/c^2 可看作是零。但是，如果 v_1 和 v_2 中的一个(或甚至是二者)等于 c，奇怪的事

就开始发生了。你绝不可能把两个小于光速的速度相加而得到一个大于光速的结果。

类似的方程式(这里不给出具体数学式子)告诉我们，当运动物体相对于我们所在的参考系的速度趋近光速时，其质量会变大，同时该物体在其运动方向上的长度会缩短。一只运动的钟与在我们的参考系里静止的钟相比会走得更慢。更有趣的是，两个事件同时发生这一概念只在一个参考系里才有意义，一个以恒定速度相对于你运动的观测者对于事件的先后或同时发生有着与你不同的看法。所有这些在今天的工程学中都起作用。能把诸如质子和电子那样的粒子加速到接近光速的机器是依据爱因斯坦的方程式来建造的。假如那些方程式没有很好地描述世界运行的方式，那些机器也就不会运行；而随着那些机器的运行，物理学家就能对质量增大、时间延缓以及爱因斯坦所预言的其他效应作出直接测量。狭义相对论与较老的牛顿力学(如果不涉及接近光速的运动它仍然足够准确)和麦克斯韦提出的电磁学方程合在一起，对日常世界作出了极好的描述。但是，它仍然只是“狭义”相对论。它不适宜于处理引力，而引力是支配宇宙整体的力。所以它不能对宇宙整体作完备的描述。要这样做，爱因斯坦就需要一个更普遍的理论。

狭义相对论是它那个时代的产物。那时明显地需要调和牛顿理论和麦克斯韦理论，假如爱因斯坦没有在 1905 年得出狭义相对论，可能晚上一两年别的什么人也会做到。但是广义相对论不同。没有人想到爱因斯坦会对狭义相对论的局限性如此不安。他那一代人中除他之外恐怕没有人能够提出广义相对论。但在 10 年以后(并不是一心只对付这一个难题，他在那 10 年中还对量子理论作出了其他重要贡献)，他创造出了一个比对宇宙的已有观测远为完整的理论。那时的观测家们还没有确定星云的距离尺度，还不能肯定星云就是别的星系，更不

用说那些星系几乎全都在高速退行了，爱因斯坦却创造出了一种理论，从中能够自然地、几乎是自动地得出这样一个宇宙：它很大很空旷，并且应该在膨胀。爱因斯坦并没有试图用他的方程来描述我们这个宇宙。他的主要兴趣是得到一个宇宙模型——一个数学模型，以便核实广义相对论的确能够描述完整的宇宙，而不涉及在无限远处或在宇宙“边缘”的所谓边界条件问题。这种模型具有深刻得多的数学真实性。所以他并不很在乎他的简单、完整、自洽、不需要特别的边界条件的理论似乎不能描述实际的宇宙。他更关心的是他的模型的确完整，并且不需要特别的边界条件。在一定意义上讲，他没有接受自己的理论所要告诉自己的东西。他一生中这一次没有遵循抛弃一切成见的准则。为了使他的模型能更符合他所抱有的宇宙是静态的这一先入之见，他把自己的方程作了一点改动，变得稍为复杂一点，以得出一个稍稍不同的无特殊边界条件的完整模型。

广义相对论首先是一种引力理论。几乎正好在完成狭义相对论与发表广义相对论之间的半路上，即 1911 年，爱因斯坦在《物理学杂志》上发表了另一篇论文，从中可看出他正在如何构思一种引力理论。该论文的题目是“引力对光传播的影响”，尽管其中内容是半真实半猜想的混合而没有任何炫目的新见解，却指出了前进的道路并给出了又一个能深刻揭示宇宙之真实而又很朴素的方程。爱因斯坦对引力会被物体的下落运动所抵消这一事实印象极深，这种抵消不是只对一个下落物体而言，而是对所有下落物体都同样如此。伽利略已经指出，所有物体都以同样速度下落，而不论其质量多大；牛顿已经运用了这一洞察来建立自己的运动定律。力对物体的作用是产生加速度，产生同样加速度所需力的大小正比于物体的质量，这是著名的牛顿三定律之一。作用在一个物体上的引力的大小也正比于物体质量。于是

质量可以消去，所有物体都以同样速度下落。

爱因斯坦在关注那个跟着光线跑的人之后，看来又用了很多时间去关注另一个（或许是同一个？）关在断了缆绳而下落的电梯里的人。这是爱因斯坦思考在引力作用下自由下落物体之行为的奇妙方式。在下落的电梯里，所有东西都以同样的速度下落，没有相对运动。电梯里的人将在空中漂浮，完全失重，毫不费力地在电梯的四壁和上下板之间游荡。当然，我们现在已经看到宇航员在飞船里这样做的图像，他们能这样做也是由于同样道理，他们也在引力作用下“自由下落”，因为一个环绕地球的轨道也是一种特殊的受控制的下落。但是爱因斯坦不得不想象我们已在电视上看到的一切：在下落的电梯里铅笔失重而浮在半空，水不是往下流而是形成圆球，如此等等。下落电梯（或飞船）里的物体遵循我们在中学里就学过的牛顿定律——不受力的物体保持匀速直线运动。在电梯外面的世界里，事情会由于引力而有所不同。爱因斯坦的天才就在于能看到别人都忽略的要点。既然下落电梯的加速度能够精确地抵消掉引力，这就表明引力和加速度必定精确地相等。

为什么这一见解非常重要呢？设想把电梯代之以一个没有开窗户的大实验室。通常的实验室是建在地球表面上，其中的物理学家可以测量物体如何下落，并从而求出引力。现在想象实验室是在空中，其中的物理学家也没有什么困难就能知道自己在自由下落。但是如果实验室受到一个恒定的、大小与在地球表面上的引力精确相等的、朝向“上方”（借助于实验室的地板和天花板来判断）的力推动，又会怎么样呢？实验室里的所有物体都朝向地板下落，正像飞机里的乘客在飞机起飞时被压到座椅背上一样。乘客感受到的压力很快就随着飞机以匀速飞行而消失。但只要那个实验室在向上加速，那向下的力就会持续。物理学家可以重做所有实验并得到与在地面上的静止实验室里完

全一样的结果。没有任何办法可以区分实验室是在一个引力场中静止，还是在被向上加速。引力与加速度是等效的。

这些对光而言又会怎样呢？回到那个被恒力推动的空中实验室。[8]其中的物理学家可能准备做些涉及光的实验。他让一道光束从实验室的一端发出并横贯到另一端。光越过实验室需要一定时间，在这段时间里实验室被持续地向上加速，所以室壁在光到达之前已经向上移动。[9]物理学家原则上可以由光到达墙壁的位置下移得知实验室是在加速，他甚至可以测量光束的弯曲程度来算出加速度有多大。这样看来，终究有了一个区分引力和加速的办法。但是爱因斯坦说，哪儿的话！我们必须保持引力与加速度等效的思想，直到（或除非）能证明它们并不等效。如果光束在一个加速参考系里是弯曲的，那么，要是理论正确的话，则光束也必定被引力弯曲，并且弯曲程度完全相等。

这个等效原理是1911年论文中正确洞察的核心。[10]不幸的是，文中对光弯曲程度的定量计算错了，但这并不要紧。4年后爱因斯坦建立了完整的广义相对论。这个完整的理论也预言光会被引力弯曲，弯曲量比1911年计算的要大。理解这种弯曲如何发生的最好途径，是抛弃我们对于力和空间的先入之见而跟上爱因斯坦在1915年初步地、在1916年以完整形式表述的思想。那就是，我们日常认为空无一物的空间，其实是一种几乎有形的、有4个维度（空间3维和时间1维）的连续体，它会被存在于其中的实物所弯曲和变形。这种弯曲和变形就表现为“引力”。

暂且放下4维时空而考虑一个有弹性的2维面。想象一块橡皮布被紧绷在一个架子上成为一个平面，这就是一个爱因斯坦版本的真空“模型”。现在有一个很重的保龄球落到了布的中央，布就弯曲了。这就是爱因斯坦的大量物质附近的空间变形的“模型”。你让一个小弹子在平直橡皮布上滚动时，它的路线是直线。但当布已因保龄球而

变形时，小弹子在球附近的滚动路线就会由于布的凹陷而变成曲线。这就是爱因斯坦所解释的“引力”由来。其实没有任何力，物体只是简单地沿着一条最小阻尼的路线，一条等价于直线的路线，来通过空间或时空中的一个弯曲部分。这里所说的物体可以是小弹子，是行星，或是光线，而效应都是一样的。当物体在大的质量附近运动时，或者按照老的图像来说通过一个引力场时，运动路径会弯曲。广义相对论精确地预言了光线在经过太阳附近时将弯曲多少。有关的数学也许很深奥，而像弯曲空间这样的概念也的确稀奇古怪。但是爱因斯坦的理论给出了明确的、可检验的预言。那是 1916 年，爱因斯坦在德国工作，英国天文学家爱丁顿从中立国荷兰的一位同行那里得知了这个新理论及其预言。德国人的预言由英国人在 1919 年作的观测所证实，那时这两个国家从法律上讲还处于交战状态，因为签订的只是停战协定而不是和平条约。部分地由于这种历史背景，1919 年的观测验证在公众中引起了其他物理学发现未曾有过的轰动，唯有 19 世纪达尔文进化论的发表可与之相比。

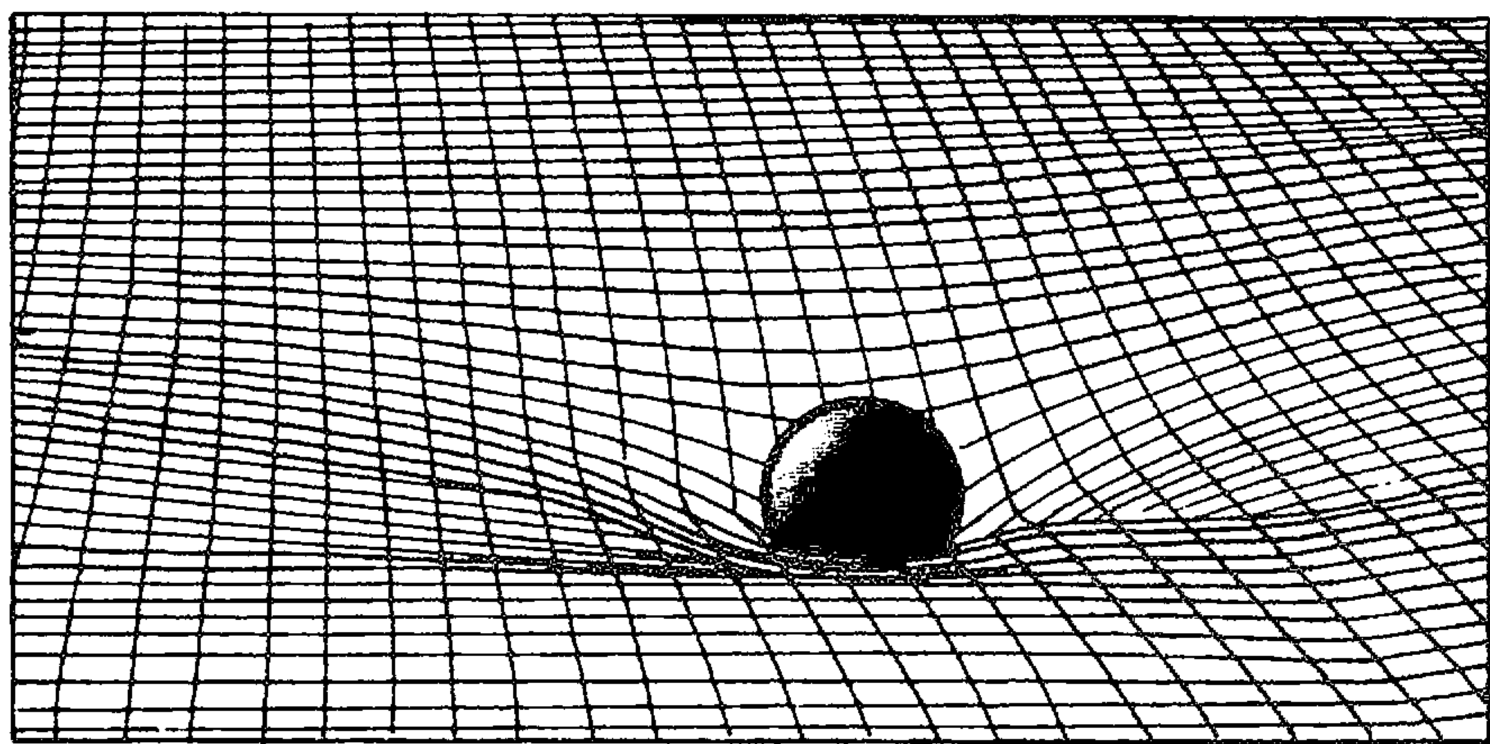


图 3.2 物质使其附近的时空发生弯曲，这可以用重球使平直的橡皮布弯曲来比拟。

验证

严格说来，科学理论是不可能被证明为正确的。任何理论家的最

好希望莫过于，他或她的理论所作的预言可付诸检验，并且在观测或实验的误差范围内是精确的。在这个意义上讲，爱因斯坦的理论已被证明为是比牛顿引力理论更完整的理论，因为它的预言与观测符合得更好。正是在这个特定的意义上，爱因斯坦理论在 1919 年被“证明”为正确。对这一验证作出主要贡献的人就是英国天文学家爱丁顿。

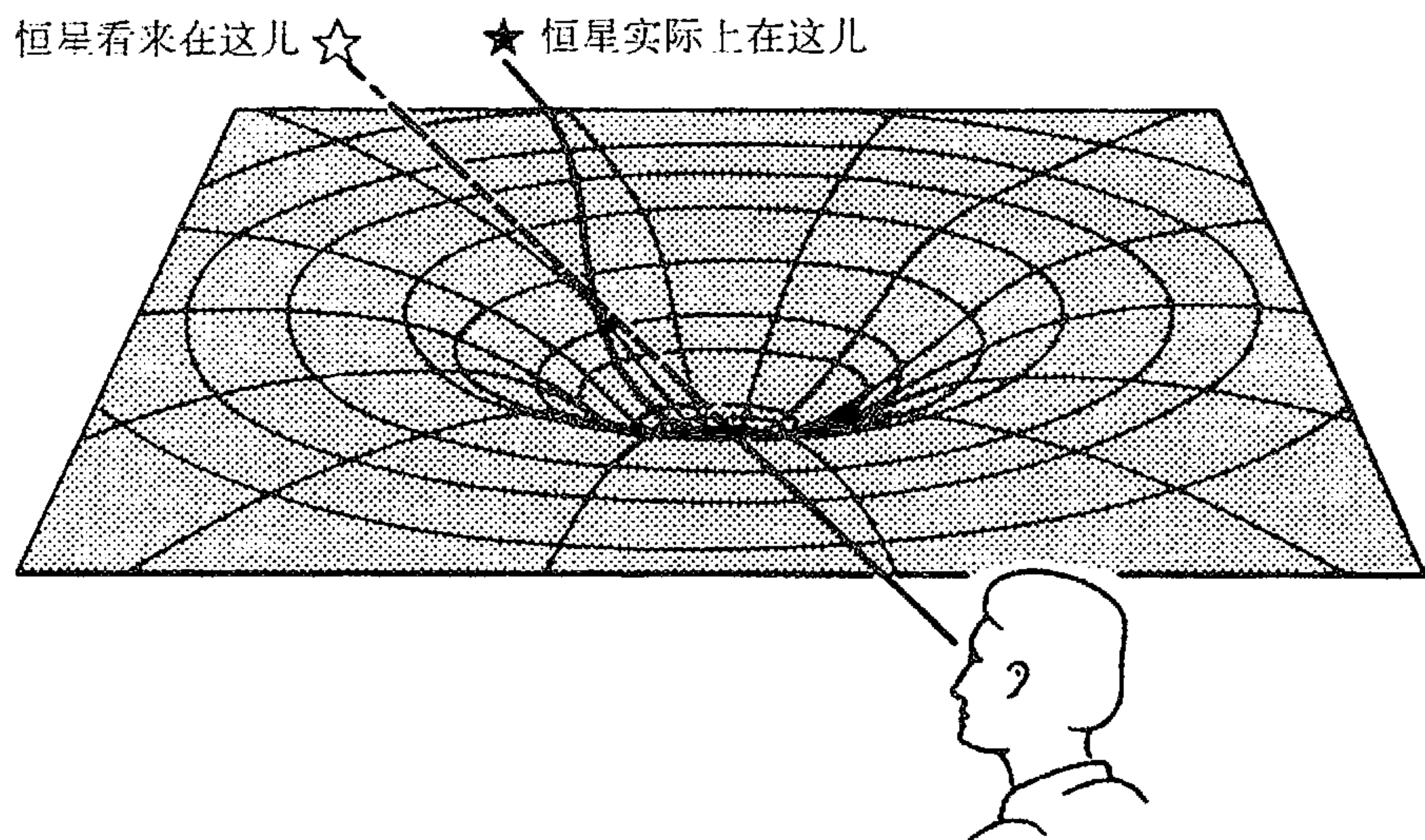


图 3.3 光在时空中的运动是沿着由物质造成的弯曲路线。重球压弯橡皮布也可以用来显示太阳造成的时空弯曲，以及这种弯曲对来自遥远恒星的光的效应。

爱丁顿比爱因斯坦小 3 岁，1882 年生于英格兰坎伯里亚的肯德尔。但是他父亲在 1884 年就去世了，爱丁顿随母亲和姐姐移居到英格兰萨默塞特的滨海韦尔顿，后来在那里上学。他终身是一名贵格会教徒，这一点对证实爱因斯坦关于光线弯曲的预言很重要，不过是以一种迂回的方式。他是一位杰出的学者，先入读曼彻斯特的欧文学院（曼彻斯特大学的前身），1902 年毕业后来到剑桥，3 年后毕业于剑桥大学，教了一段书后于 1907 年成为三一学院的研究员和皇家格林尼治天文台的首席助理，1912 年 30 岁时当上剑桥大学天文学和实验哲学（我特别喜欢这个名称）的普鲁密安教授，1914 年就任剑桥天文台台长。

如果这一切使他听起来令人生畏，那么这个印象只是半对。他还有卓越的表述才能，是20世纪20年代和30年代最主要的科学普及者之一。他很有幽默感，也有点古怪。他在晚年曾讲过自己小时候玩过的一个学童游戏，就是造出一个短句完全符合语法却毫无意义，例如“站在树篱旁，听来像萝卜”。他介绍量子物理学和相对论等理论的著作里，常喜欢引用刘易斯·卡罗尔(Lewis Carroll)的作品来说明要点。一个能在一本题为《物理科学的哲学》的书中用下面这句话作为一章开头的人，必定有什么超越学术常规之处：

> 我相信宇宙中有15 747 724 136 275 002 577 605 653 961 181 555 468 044 717 914 527 116 709 366 231 425 076 185 631 031 296个质子和相同数目的电子。

如我们将在下文看到的，或许更奇怪的是爱丁顿怎么得出这个大数字的缘由仍在吸引着宇宙学家们。

爱丁顿将以其两大成就而被永久纪念。他对创立天体物理学的贡献不亚于其他任何人，该学科研究的是，如何将在地球上得出的物理定律与对星光的观测结合起来，去解释恒星内部进行着的使高温状态得以保持的过程，并探索恒星如何随其年龄而变化。他又是用英语普及爱因斯坦相对论的权威，不仅是把那些思想传达给外行读者，而且作出科学的解释使之清楚地呈现在他的同行面前，还撰写专门的教科书来推动其流传。尽管爱丁顿的整个一生和工作都富有魅力，我只能在这里举出一件事，就是他对光线经过太阳附近时必定被弯曲这一预言的反应。

爱因斯坦关于广义相对论的首篇论文于1915年下半年寄到柏林

科学院，并以更详细的形式于次年发表。[11]该论文的副本很自然就到了他在中立国荷兰的朋友手里，朋友中有一位即德西特(Willem de Sitter)，则寄了一份副本给爱丁顿。在1916年和1917年，德西特还将自己的3篇文章寄给皇家天文学会以求发表。这些文章部分地是对爱因斯坦工作的介绍和对其意义的诠释，但在第三篇里德西特还首次提出了一个以广义相对论为基础的宇宙模型，按照这个模型宇宙应该膨胀，这还将在下文谈到。爱丁顿那时是皇家天文学会的秘书，我们知道他仔细阅读了那些文章并在它们发表之前在学会的会议上作了介绍。这位有智力和知识背景来充分地鉴赏爱因斯坦新工作的意义的人，正好在恰当的地方和恰当的时间得到了消息。但是爱因斯坦的新理论在被证明为正确之前还有几次命中注定的曲折。

爱因斯坦所提出的检验光线弯曲的方法是在日食时观测太阳附近的恒星。在通常情况下，太阳的光芒当然会使得它所在那部分天空的恒星不可能被看到；但当太阳暂时被月亮遮暗时，就可以拍摄下那些比太阳远得多但与太阳在天空中同一方向上的恒星的位置。把这些照片与6个月之前或之后(即太阳在地球的另一侧时)对同一块天空拍摄的照片作比较，就有可能看出由光线弯曲效应造成的恒星视位置的移动。天文学家需要的是一次日食。假如他们能挑选日食的时间，那就会要任何一年的5月29日，因为那时太阳会从毕星团方向上一块亮星密布的区域前经过。从地球上的不同地方倒是常能看到日食，但是5月29日(或者一年中任何其他特定日子)的日食就很罕见。如爱丁顿自己所说的，“有可能要等上几千年才会在这个吉日发生一次日全食。”但是命运之神真是青睐，1919年就有一次日食，而且正是在5月29日。机会太好了，绝不能错过，而且连第一次世界大战都似乎是为便于组织观测这次日食的远征而及时地结束了，能看到日食的地

方是巴西和非洲西海岸外的普林西比岛。

但在1917年，故事情节开始变得复杂起来。皇家天文学家弗兰克·戴森爵士(Sir Frank Dyson)热烈支持组织两支远征队去观测1919年日食，并紧急制订了计划。与此同时，英国正在征兵，所有体格健全的男人都要准备入伍。爱丁顿那时34岁，身体很好；但他还是一名虔诚的贵格会教徒，并因此而拒服兵役。这在1917年可是一件麻烦事，而当科学界识别出他是一位一流学者时，情况就更为复杂了。物理学家仍对X射线结晶学的先驱者莫塞莱(Henry Moseley)于1915年在加利波利的战斗中阵亡记忆犹新，并质问政府为什么要把最好的科学家送去死在战场上。一群知名学者敦促内政部给予爱丁顿豁免，理由是让他继续自己的适当工作最符合英国的长远利益。内政部终于同意了，并寄给爱丁顿一份公函让他签字后寄回。但是爱丁顿在公函上加了一个脚注，说是如果他不是以所述理由而缓役，他将以良心为理由要求缓役。这是一个诚实的有原则的立场，它给了内政部一个难题，更使那些为他辩护的学者不安。按照当时的法律，拒服兵役者要被送去干那些不会惬意的农业或工业活，而爱丁顿已经准备好去和他的贵格会朋友们一起干那些活。又是一番争论，戴森作为皇家天文学家也参与了，最后结果是，爱丁顿的兵役可以延缓，但有一个“条件”，就是如果战争到1919年5月已结束，他就必须率领一个远征队去检验爱因斯坦关于光线弯曲的理论预言！[12]

爱丁顿曾于1912年率队去巴西观测日食。他需要运用自己的全部经验来保证1919年两支远征队中他这一部分的成功，整个1918年都在进行准备工作。计划是，爱丁顿和剑桥队去普林西比，而戴森从格林尼治皇家天文台组织一支队伍去巴西。但是在停战协定签署之前仪器制造工作无法进行，工厂都忙于制造战争武器。远征队必须在

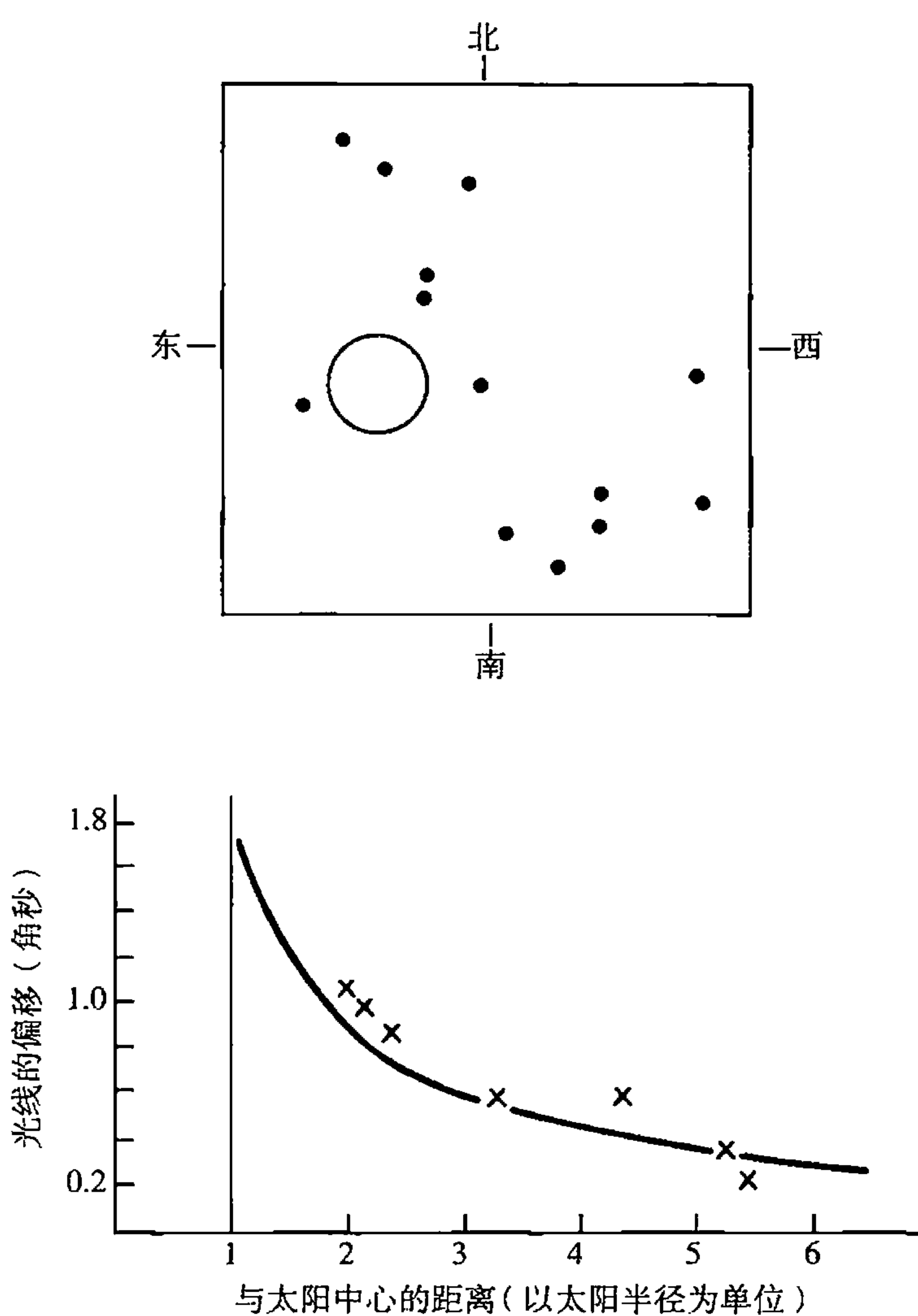


图 3.4　爱丁顿在 1919 年日食时观测到的一些恒星位置如上图所示，当时那些恒星和太阳（图中圆圈）近乎在天空中的同一方向。恒星距离我们当然比太阳要远得多，这种并列就意味着来自遥远恒星的光经过了受太阳引力影响的空间区域，如图 3.3 所示。

当这些恒星的位置与太阳位于天空中相反一侧时测到的同一些恒星的位置比较时，爱丁顿发现了明显的偏移，各颗恒星位置的偏移量取决于日食时该星与太阳的角距离。光在经过太阳附近时被弯曲了。恒星位置的偏移（下图中用 × 号表示）与爱因斯坦理论所预言的（下图中曲线）精确相等。

1919 年 2 月起航，而停战协定是在 1918 年 11 月 11 日签署。在那么几个闹哄哄的星期里一切都准备就绪，远征队出发了。巴西支队遇上了极好的天气，拍到了一套效果极佳的日食时太阳周围星场的照相底片。但是由于后勤方面的原因，那些底片没有被马上冲洗和研究。而

爱丁顿却在普林西比焦急地等待着，那天开始时在下雨，天空阴云密布。所有拍摄的准备工作都已做好，只能盼望而不敢预期了。就在接近全食的时候，太阳隐约地出现了，于是赶紧拍摄。结果只有两张底片上显示出检验所需要的恒星。爱丁顿已经安排好对拍到的底片当场检测，如他自己所说："不完全是出于急切，也是怕在返程中有什么意外。"那两张底片中的一张就在普林西比被及时地冲洗出来并作了分析，爱丁顿把它与自己带来的同一块天空的另一张照片作了比较。有关的测量是很简单的。日食之后3天，爱丁顿知道自己手中已经有了证明爱因斯坦广义相对论正确的证据。

日食观测的完整分析又用了几个月，爱因斯坦得知自己的预言被证实的确切消息已是1919年9月。整个观测结果是在1919年11月6日皇家学会和皇家天文学会联合举行的一次座无虚席的会议上宣布的，当时那个渴望除了战争以外的任何新消息的世界立即出现了一股宣传浪潮。报刊的大字标题是"光不走直线"、"科学的革命"、"牛顿理论被推翻了"、"空间'弯曲'"等等。爱因斯坦在公众眼里成了20世纪或许是有史以来最伟大的科学家。而广义相对论则被认为是有史以来最伟大的科学理论——这恐怕不大对，因为量子理论至少是同等重要。爱因斯坦的理论还有其他的验证。它在此之前已经解释了牛顿引力理论所不能解释的水星绕日轨道的一种微小变化，所以可以说日食结果只是证实了天文学家已经知道的事情，即爱因斯坦的理论比牛顿的好。后来又有其他的日食观测，[13]检验工作已重复了许多次，并且比爱丁顿在普林西比所做的要精确得多。对广义相对论的其他不同途径的检验，包括星光由引力造成的红移和脉冲星辐射的细微变化(后一项在1919年时是怎么也想不到的)，全都给出同样的结论。但是，不论爱因斯坦理论在1919年前已成功地解释过什么，也

不论后来又作过些什么检验，1919 年 5 月 29 日才是科学观测证明这一理论正确的日子，1919 年 11 月 6 日才是公众知道这一事实的日子。不过，天文学家还是有点不解：如果爱因斯坦理论对时间和空间作了这么好的描述，为什么它只涉及宇宙中一些很特殊的现象呢？

爱因斯坦的宇宙

广义相对论是关于宇宙的几何亦即时空几何的理论。在过去 50 年或更长时间里，最熟悉广义相对论的方程和意义的一位宇宙学家是萨塞克斯大学的麦克雷(Bill McCrea)，现在以荣誉教授和麦克雷爵士的头衔依然活跃。他 1904 年出生，1926 年毕业于剑桥大学三一学院，正好属于在学术生涯开始时首先接受爱因斯坦新思想的那一代。他在自己那漫长而杰出的生涯早期，曾和米尔恩(Edward Arthur Milne)一起试图证明，即使用牛顿引力理论，只要再加上一定的简化假设，也可以作出与广义相对论类似的关于宇宙演化，尤其是其膨胀的预言。他还研究过星系演化，琢磨过那些使爱丁顿着迷的巨大数字的意义，并对量子物理学和恒星天文学作出了重要贡献。在当代没有别人能像他那样对广义相对论的含义作出如此明晰的解释，我的宇宙学知识也是在 20 世纪 60 年代后期当他的学生时学来的。[14] 所以，我不再沿着那引至对广义相对论的现代认识的历史进程蹒跚，并且向爱因斯坦和爱丁顿表示适当歉意，我将在这里给出我自己对广义相对论的解释。

狭义相对论用某种特别的数学描述把时间和空间结合成一个物理本体，即时空。用数学语言讲，这种时空是几何上“平直”的，它与地板面或台球桌面有着同一类型的几何。所以它是一个包含更广泛可能性的族即曲面族的一种特殊情况。对数学家来说，“曲”面是指任

何非平直的面，如地面上起伏的山峰和山谷，还有地球这个近似球体的表面本身。一个严格的球面也和严格的平面一样是特殊情况，而曲面上也可以有小起伏，正如地面上有山峰和山谷。

类似地，时空也可以是弯曲的，正是这一推广导致了广义相对论。宇宙学家在试图从数学上描述(或预言)宇宙的行为时，从众多可能性中选取一两种曲面来看看是否能最好地与观测相符。被选取的范例就叫做模型，但它们并不具有雕塑模型那样的物理真实性，它们仅仅存在于宇宙学家的头脑里和方程式中。这样的模型只解释宇宙的宏观特征，而小的起伏就好比地面上的峰谷，尺度(如果称得上尺度的话)小得不足以纳入计算。太阳使其周围时空弯曲，从而使经过其附近的光线弯曲的效应，在宇宙这个大格局里正是一个极细微的起伏。

广义相对论在另一个意义上也是狭义相对论的发展。如我已讲过的，它既描述物质又描述时间和空间。狭义相对论给了我们时空，广义相对论则实际上给了我们"物质时空"，尽管我从未见过谁使用这么一个词。正是物质使时空网弯曲或曰畸变，而广义相对论为一个完全确定的物质、时间和空间的几何给出了明确的物理意义，那就是一个宇宙，这里所说的宇宙是指许多可能的数学模型中的一个，不一定是我们生活于其中的这个真实宇宙。严格说来，广义相对论只适用于那些完整的宇宙。当爱因斯坦方程被运用来描述太阳附近光线的弯曲或水星绕太阳旋转轨道的微小移动时，这种运用只是在近似的意义上。实际上，这种近似的误差可以做到要多小就有多小。描述像太阳这样的小小局域物体的方程可以由所谓的"边界条件"而连接到宇宙的其余部分。但重要的是，爱因斯坦并不需要推广他的理论使之能处理整个宇宙。广义相对论从它诞生之时起就能很舒服地对付所有那些完整的宇宙，让它缩在像太阳系这样小小的、不重要的时空范围里才

真是大材小用了。

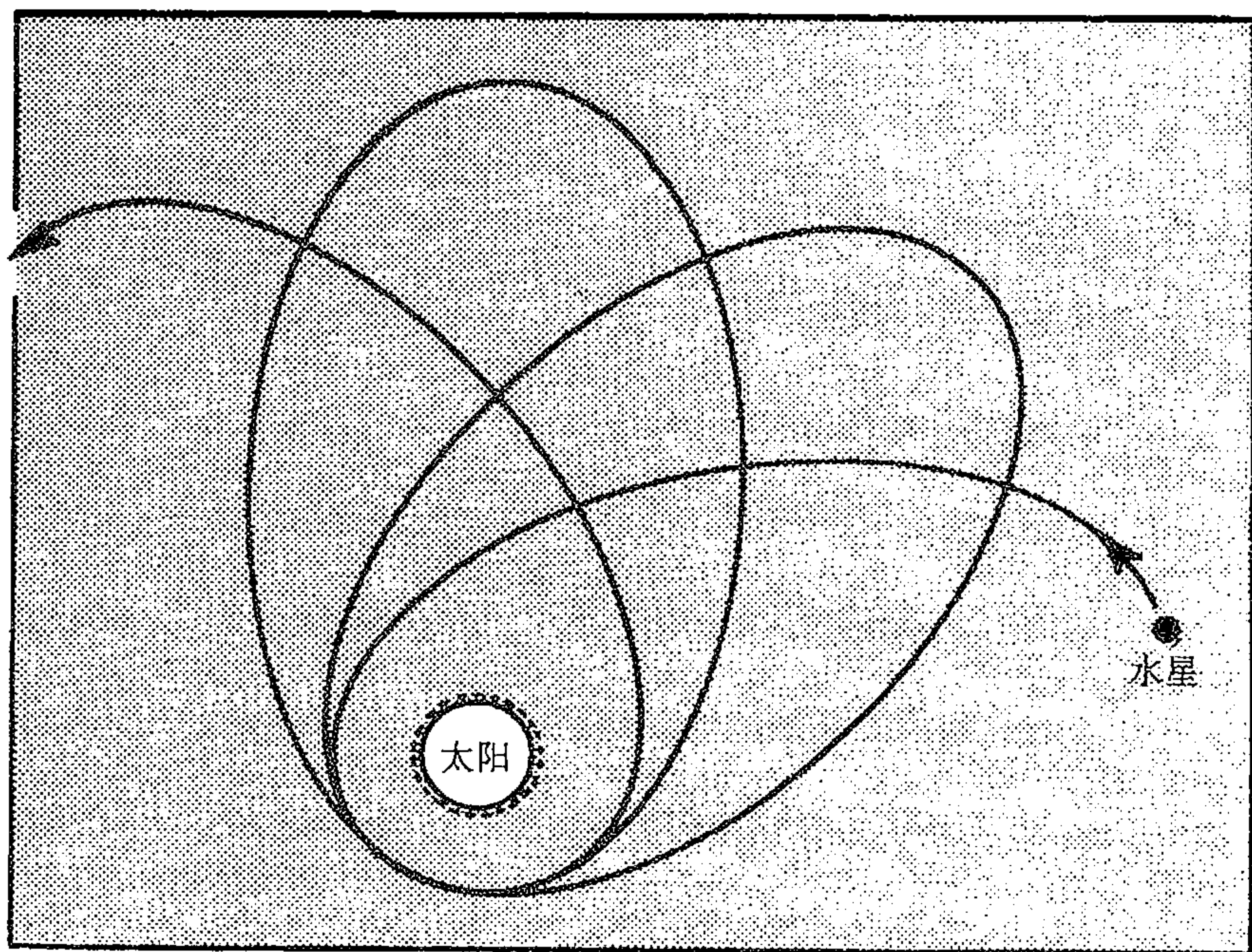


图 3.5　水星沿椭圆形轨道绕太阳旋转，而整个轨道也在转动，形如花瓣。牛顿引力理论不能说明这种轨道转动，广义相对论则能很好地予以解释。

1917 年的标准认识仍是，银河系即是整个宇宙，宇宙也就是恒星的稳定集合。单个的恒星可以四处漫游，但若把银河系作为一个整体看，则宇宙的突出特征似乎就是稳定性。银河系既没有变大也没有变小，它就在那里，并将永远如此。有的恒星在诞生，有的在死亡，但银河系的整体面貌总保持稳态不变。所以当爱因斯坦拿起自己发明的新奇工具用于描述宇宙时，他预期的结果是至少有允许稳态宇宙存在的可能性。在 1917 年寄给柏林科学院的另一篇文章里，他描绘了自己对方程式所给出结果的惊讶，以及如何设法迫使那些方程式适合他先入之见的框框。该文的标题是“基于广义相对论的宇宙学考虑”，麦克雷引用了爱因斯坦在文中的话(经翻译)：“我将引导读者走上我已经走过的这条颇为崎岖曲折的道路，因为否则我就不能指望他会对最后的结果感兴趣。”麦克雷称整篇文章是“试验性的……缺乏爱因

斯坦的特色”。[15]的确，这位伟大人物对自己的劳动果实感到困惑了，这是因为，如我们现在知道的，那时对宇宙的观测给了人们错误的印象。

爱因斯坦试图描绘一个最简单而又与真实宇宙有关的可能宇宙模型，即物质在空间均匀散布的宇宙。这个宇宙是封闭的，像一个球的表面那样自我弯曲闭合，因而没有边界。[16]但是这个宇宙将不会保持静止。当爱因斯坦把他的广义相对论方程运用到这样一个宇宙时，发现它必定要么膨胀要么收缩，而不可能保持静止。他能使模型宇宙静止从而与银河系的表象相符的唯一办法，就是在广义相对论方程中额外增加一项，这一项叫做“宇宙常数”，常用希腊字母 λ 表示。这样，也只有这样，方程才能描绘出一个既不膨胀也不收缩的宇宙。爱因斯坦 1917 年文章的最后一句是这样写的：“需要这个补充项只是为了使物质的准静态分布成为可能，而这种分布是同恒星的速度很小这一事实相符的。”十几年后，哈勃揭示出银河系并不是整个宇宙，并且发现了遥远星系在以巨大速度退行。引入 λ 项的理由原来并不存在，而膨胀宇宙却成了爱因斯坦本可作出而未能作出的最重大预言。爱因斯坦的方程本来要告诉他自己的是，我们所生活的宇宙是动态的膨胀的而不是静态的，这对所有天文学家来说已十分清楚。爱因斯坦后来把引入 λ 项称为“我一生中的最大错误”，但是也很难要求任何一个人在 1917 年研究宇宙的性质时能不犯这样的错误。理论本已走在观测前面，但只有当哈勃和赫马森发现我们确实生活在一个膨胀的宇宙即爱因斯坦的宇宙中时，广义相对论才终于取得了它描述整个宇宙的起源和演化的应有地位。

但在详细讲解对宇宙的这种描述亦即大爆炸模型本身之前，也许值得先暂停一下，来看看另一种也很惊人的途径，它能从很简单的观

测出发对宇宙的性质作出非常有力的推论。不管银河系给人的表面印象如何，你并不是一定要有了对遥远星系的观测才能知道宇宙必定处在变化的状态，而如果爱因斯坦对前辈哲学家的某些思想更熟悉一些的话，他恐怕就不会陷入 λ 项的混乱。另一方面，他对宇宙中物质所起作用的关注，部分地来自另一条半哲学的思路，那条路可以回溯到牛顿，即使在 1917 年都可用来预言哈勃关于宇宙中物质分布的某些发现。而这两条哲学思路本可合起来作出宇宙中布满星系、星系在互相散开的预言，但是这个预言从来没有作出过。

注　释：

1. 这方面还有后续的故事，20 世纪 80 年代的新突破使天文摄影家能得到天体最好的像和最好的光谱。马林(David Malin)和墨丁(Paul Murdin)在他们的书《恒星的颜色》(*Colours of the Stars*)中介绍了许多这方面的技术并展示了他们拍摄的大量照片，该书由剑桥大学出版社于 1984 年出版。

2. 曾被席阿玛(Dennis Sciama)引用，未说明出处，见 *Modern Cosmology*，p.43。

3. 当然这是有缘故的，我在这里的讽刺并不真是对理论家作出公平的评判。他们正确地认为那些早期的、简单的宇宙学模型只是尝试性的，因为那些模型显然没有精密到足以说明所观测宇宙的复杂性。经过了几十年时间才得以逐渐看出，为什么那些看似儿童玩具的模型能对宇宙的本质作出合理的揭示。相信这些简单模型的理由现在已变得高度复杂，但是一个重要的基本事实仍然是，宇宙所遵循的规则竟是如此简单，以至于理论家起初认为那些规则不可能是对真实的恰当描述。

4. 例如见霍夫曼(Banesh Hoffman)著《爱因斯坦》(*Albert Einstein*)。

5. 见我的《薛定谔之猫探秘》(*In Search of Schrödinger's Cat*)一书中关于量子物理学由来的故事；我还在《空间弯曲》(*Spacewarps*)和《时间弯曲》(*Timewarps*)两书中更详细地介绍了相对论的一些非宇宙学应用。

6. 仙女座星系光的蓝移表明该星系在以很大的速度朝向我们运动，但它的光相对于我们的速度仍然只是 c。

7. 如果某一个河外星系里的一颗行星上有智慧生物，而该星系又在以 1/3 光速离开我们，那么不仅它们测量从自己星系发出的光的速度是 c，我们测量那些光到达时的速度也是 c，而不是 $2/3c$。这一切对今天的科学研究仍然不无教益。如麦克雷在评论这一章时所指出的，牛顿的思想在那些从未接触过数学物理学的人看来也会是奇怪而违反“常识”的。牛顿发明了数学物理学，这在他那个时代是革命性的。爱因斯坦发挥了牛顿提出的基本思想，但所用的也是牛顿发明的技巧。所以牛顿或许是更伟大的创新者。那么教益是什么呢？不要把任何东西看成是“常识”或“显然的”，而是用开放的头脑，不带任何先入

之见地去探讨一切，这才是爱因斯坦所做的，这才是爱因斯坦的成就背后的天才。

8. 其实不一定要恒力，为简单起见暂且如此。

9. 这里区分速度的效应和加速度的效应是很重要的。如果想象中的空中实验室以一个恒定速度运动，任何从室内一端横贯到另一端的物体也就具有那个速度，实验室和该物体就会在该速度方向有同样的位移。正是加速度，即速度的改变，使实验室在光束穿越时有一个额外的位移。但在下落的电梯里，光束也受到引力作用因而也被加速，所以恰当的理想实验场所其实应该是被某种外力（比如说火箭）推动而不是在引力作用下下落的空中实验室。

10. 麦克雷说他怀疑爱因斯坦真的是从等效原理出发来导出他的理论体系。他猜想那只是爱因斯坦在得出他的体系后所采用的一种方便的阐述方式。但是，麦克雷又说他的大多数同事不会同意他的这种说法，所以我在这里还是采用故事的传统版本。

11. 当然又是刊登于《物理学杂志》（*Annalen der Physik*，第 49 卷第 769 页）上，题为"广义相对论的基础"，被公认为是爱因斯坦一生中最重要的论文。

12. 这个故事有几种版本，侧重点稍有不同，但讲的是同样一串事件，我采用的是钱德拉塞卡《爱丁顿》（*Eddington*）一书中的叙述，该书也是本书对爱丁顿的直接和有关引文的出处。

13. 继 1919 年的两队观测之后，利克天文台的队伍到澳大利亚西部的沃勒尔观测了 1922 年的日食，作了对爱因斯坦光弯曲效应的第三次测量并给出了广义相对论所预言的结果。这个消息在 1923 年 4 月的皇家天文学会会议上宣布时，爱丁顿又先搬出刘易斯·卡罗尔，说了一番很显其个性的话："我记得《狩猎蛇鲨》里的贝尔曼制定了一条规则'我说三次，就是对的'，现在那些恒星已经分别对三个观测队说了三次，所以我相信是对的。"

14. 麦克雷为我所编的《今日宇宙学》（*Cosmology Today*）一书提供的文章，对广义相对论作了很好而又易读的解说。若非有他评阅本章的话我是不敢冒险在这里讲述广义相对论的，他说我的讲述是"相当能自圆其说的"，我想这就是说还行，如果不全符合他自己来讲这个故事时所会采用的方式的话。

15. 引自麦克雷，在前引文中。

16. 米尔恩和麦克雷所提出的"牛顿宇宙学"不能妥善处理边界问题，因为它不允许时空弯曲，它也就不能真正对我们的宇宙或任何宇宙作出适当的描述。所以在任何实际意义上都不能说它是相对论宇宙学的竞争对手，不过它还是有其用处。

插曲一　哲学家的宇宙

对于一个从屋顶上自由落下的观测者来说引力场是不存在的。

——爱因斯坦

爱因斯坦给了我们广义相对论这个数学工具，用它可以构造关于我们居于其中的宇宙的详细模型。这些模型使宇宙学家几乎违心地被引至宇宙产生于一种超密态亦即大爆炸的思想。广义相对论把宇宙学从哲学家和形而上学家的领域取出来，并使之成为科学的一部分。但是，在我们细看对宇宙学的科学研究的含义之前，不妨从事后之见的优越角度来考虑一下，如果哲学家只凭自己由信念而产生的勇气，他们在同一条路径上究竟有可能走多远。

爱因斯坦在建立广义相对论的时候，曾为他从奥地利物理学家马赫那里得知的一个思想深深吸引。他把这个思想称为“马赫原理”，尽管事实上它可以追溯到比马赫更早得多的时候。该思想把很大的事物——宇宙整体，同很小的事物——地球上实验室里或日常生活中的

物体联系起来。爱因斯坦本打算把马赫原理吸收进广义相对论，把宇宙整体同日常的小世界连在一起，但似乎不能做得令人满意，所以他后来整个地否决了这种意图，而具有讽刺意味的是，那位老人马赫也在1916年去世前不久否决了广义相对论。也许他们俩都有点太性急了，因为有那么几位数学家正在沿着一条在20世纪80年代及此前几十年的研究中不受重视的小道重新导出爱因斯坦的方程，并正在证实大与小之间确有联系，而广义相对论方程又正好为认识这种联系提供了框架。还有，这种新研究可以应用于一些1916年时做梦也想不到的奇异天体，如自转的脉冲星和黑洞X射线源等。那些奇异天体与对大爆炸的探索仅有着勉强够格的关联。但是，20世纪20年代时已得到表述并为人所知的马赫原理再加上一点小小的推理，就应该使哲学家仅仅依据已知的银河系结构便能预言宇宙中其他星系的存在。没有任何人，连爱因斯坦也不例外，对马赫原理重视到这样来运用。哲学家对自己的玄思冥想竟被证明真能提供对真实宇宙的新知识而大为惊讶，而宇宙学家，即科学家中最哲学化的那一些，却似乎继承了那种对自己的工作总不大相信的态度。所以我们不妨停留一会儿，不必太认真地看看一些在20世纪20年代初本该就有的东西。

惯性难题

当你往咖啡里加进奶脂并搅动，然后悠闲地注视那优美的旋涡图案时，你可曾想一想那液体怎么“知道”它在运动？这个问题可比乍听之下要深刻。容易想到的答案是，液体“知道”它在转动是因为它在经过杯子的内侧面。但这完全不对。液体表面形状由平直到凹陷的变化，是在试图告诉我们关于宇宙本质和惯性本质的一些非常深刻的东西。

看来是伽利略第一个意识到，不是物体运动的速度而是其加速度才能显示出有力作用于其上。在地球上总是有外力，例如摩擦力，使运动物体减速，除非我们持续地推动该物体；我们能实现的最接近于无阻碍的运动是冰球在平整冰面上的滑动。那冰球看来确是以同一速率在同一方向上(即恒定速度)运动，直到它受到一个外力作用为止。牛顿给出了定量关系，即力等于质量乘以加速度，并把这关系扩展到解释行星在环绕太阳的轨道上运动。不受力的物体在其中以恒定速度运动的参考系叫做惯性系，牛顿相信必定有某个基础惯性系，即一个以某种方式由真空来定义的绝对的静止标准。他认为，物体除非是被外力加速，否则在真空中将以恒定速度运动。

但是要锁定真空可有点棘手。你不可能在真空上钉一个钉子并测量你相对于钉子的速度，从而搞清楚你是否在一个惯性系里。不论牛顿定律的实用价值如何，其哲学基础看来不大健全。但是牛顿却指望能由一个极简单的实验来确定基础惯性系，这实验不是用沿直线运动的物体，而是用转动的物体，具体地说就是一只水桶。

对这个实验描述得最好的人当然还是牛顿自己，他在1686年的《原理》一书中写道：

> 能区分绝对运动与相对运动的效应，是离开圆周运动的轴的力……如果将一只桶用长绳吊着，把桶旋转多次而使绳扭紧，再注入水，并抓住桶使它和水一道静止不动；然后突然松手，桶就会做与以前反向的旋转运动，当长绳松释时……水面起初与桶开始转动前一样是平的，但此后桶就逐渐把它的运动传递给水，使水明显地开始旋转，并一点点地离开中央向桶的边缘升起，形成一个凹面(正如我曾做过的实验一样)，当运动变得越来越快时，水也升得越

来越高。

你杯中的咖啡在你搅动时会出现同样现象。这里的关键是，液体与容器之间的相对运动并不要紧，要紧的是液体的绝对转动。你可以把一只水桶或咖啡杯小心地放在一个转盘的中心并使它们一起转动。即使在液体与容器相对静止时液面也会出现明显的下凹。液体“知道”自己在转动并且作了相应的反应。牛顿认为这就是相对于固定空间的转动。那些仍以完全不可观察的东西就不可能存在为由反对绝对空间思想的哲学家，就不得不为水桶的旋转效应寻求别的解释。他们用了30年，直到贝克莱(George Berkeley)，一个1685年出生的爱尔兰人，长大后成了哲学家、经济学家、数学家、物理学家和主教(不一定要按此次序排列)，才给出一个答案。

贝克莱认为，所有运动都是相对的，因而必须对照着什么东西来测量。“绝对空间”不可觉察，因而就不能作为这种参考物。假定宇宙中的一切物体都消失了，只有一个球体除外，那么就不可能设想这个球体还有什么运动，运动这个概念也将毫无意义。即使有两个完全光滑的球在轨道上相互围绕转动，仍然没有任何办法测量这种运动。但是“假定那有星星固着其上的天空突然被创造出来，我们就可以由球体相对于宇宙各个部分的位置来设想其运动。”[1] 贝克莱的意思是，正因为你杯中的咖啡知道自己在相对于遥远的恒星转动，所以它沿着杯子内壁上升来作显示。

他比他的时代超前了150年。虽然他的思想在18世纪得到了一定的流传，却被认为有点古怪因而被大大忽视了。那有点缥缈的遥远恒星怎么会影响地球上这实实在在的物体的运动呢？是什么魔力能越过空无一物的空间来保持对桶中水的控制呢？贝克莱的思想没有被他

的同时代人认真看待。但是火种已在，到 19 世纪后半叶终于由马赫燃成了火焰。马赫的名字作为飞行器相对于声速的速度单位而永存。

马赫于 1838 年出生于奥匈帝国的一个地方，那儿现在属于捷克。他在维也纳大学学习并于 1860 年以电学方面的工作获博士学位，在维也纳和格拉茨任教之后，于 1867 年去布拉格大学当实验物理教授，干了 28 年后又返回维也纳大学当归纳科学的历史和理论教授。尽管 1897 年的一次中风影响了他的右半身，到 1901 年从大学退休时他已当了 12 年的奥地利上议院议员，退休后他仍保持智力上的活跃，直至 1916 年 78 岁生日后一天去世。他的兴趣范围很广，从实验和理论物理学到科学哲学都在其内。他对爱因斯坦的影响是通过那本于 1883 年出版的《力学》（英译本《力学的科学》于 1919 年在芝加哥出版），该书的内容之一是他关于惯性思想的陈述。或许我们应该先回顾一下那困扰了马赫、还曾困扰了贝克莱主教的惯性力的性质，然后再看看爱因斯坦是怎样试图去解释的。

离心神话

物理学家乞求用来解释转动参考系中物体运动的惯性力实际上并不存在。它只是几种对认识上述运动有帮助的虚构力之一，那些虚构力全都源于运动物体对从直线轨道上偏离的反抗，即物体的惯性。这里的把戏是，先设想转动参考系停下来，然后设想有适当的力作用，使得物体相对于现已静止的参考系的运动就像参考系在转动并且没有虚构力时一样。

也许此类虚构力的最简单例子是科里奥利“力”，它的用处之一是能解释为什么英国的恶劣天气来自西边的大西洋远处。

由于地球在转动，赤道上的空气具有角动量，其运动自西向东，

每24小时绕地球40 035千米的圆周环行一次，也就是说其速度为每小时1668千米。两极的空气却完全不转动，没有角动量。在中纬度上的空气则处于适当的中间状态。太阳的照射使热带地区的空气温度升高，出现对流运动，在大气中形成环流。热空气上升并向北或向南运动到较高纬度处，取而代之的是沿海平面运动到赤道的冷空气。热空气冷却并下沉到赤道与极地之间某纬度的地球表面，那里原来的空气的角动量比赤道上空气的角动量要小，而来自热带的空气却携带着原有的角动量，带着对自己来处的“记忆”。这些空气在被摩擦作用减慢下来之前会试图以原有的速度朝原来的方向继续运动，于是其自西向东的运动就会超越地球在新纬度处的转动。

类似地，沿海平面往赤道流的空气就不只是朝南(对北半球而言)而且还朝西，这是因为其转动落后于低纬度处地球的自转，从东北方向吹到西南方向的所谓贸易风就是这样形成的。

从一个生活在地球上并把地面作为静止参考系的人看来，这就好像是有一种力在推动空气流动。这就是科里奥利力。它还被引入用来精确计算炮弹的飞行轨迹，以及“解释”为什么南北走向的河流总有一侧河床被冲刷得较厉害。*

你还可以用另一种方式来思考离心力。当你站在地面上时，引力(一种真实的力)把你向下拉。假如没有地面来挡住你，你将会朝地球中心自由落下。地面给了你一个力来抵消掉由引力产生的加速度。地面推你向上，你的重量推你向下，两个力大小相等而方向相反，所以你固定不动。

现在想象你自己是在一个快速旋转的圆筒里，圆筒在自由下落

---

* 原文为从北向南的铁轨有一侧磨损较厉害，似不确。——译者

(也许是在围绕地球的轨道上)。如果你与转动的圆筒壁是接触着的，那么按照牛顿的见解，你的自然运动状态将是继续在一条直线上，即与转动圆筒相切。但是你不可能做到，因为圆筒的地板(或壁)在压迫你的脚，把你由直线轨道向旁推出。结果是你被粘在转动圆筒内作圆周运动。真实的力是圆筒的壁对你的脚或你的背向内的推力。净效果是，只要你与转动圆筒的壁保持接触，你就停在原处，但在圆筒外面的观测者看来你是在沿圆形路线运动。运动方向的改变也与速率的改变一样是一种加速度，如果圆筒以恰当的速率旋转，你将感觉到与在地面上时完全一样的加速力——你若是站到转动圆筒壁上的体重计上，你会看到自己的体重与平时在浴室里称的一样。既然感觉与站在地面上房间里的地板上一样，人的大脑就很容易被骗，以为圆筒是静止的，而引力在使物体“向下”保持在圆筒地板上。但只要在那种环境里试试泼出液体或者拍一个球，真相马上就会像引入科里奥利力的情况一样明显。

除了那杯咖啡之外，关于所有这一切最为人们熟悉的事例就是科学博物馆里的大傅科摆了。这个摆(所有其他足够重的、悬挂得很好的摆都一样)开始被安装时就在空中一个平面上笨重地来回摆动，然后就保持着在同一个平面里这样摆动，尽管地球在它下面转动。而我们站在地球上却以为是摆动面在转动，这种转动的速率则取决于摆所在地方的纬度。像你杯中的咖啡一样，摆也“知道”宇宙中的所有物质是在什么地方，并使自己相对于所有遥远的星系保持静止。

在没有相对于遥远星系的转动时，科里奥利力和离心力都不存在。它们有时被归于“虚构力”，意思是不如(比如说)引力那样真实。但是这种表述再一次暴露了我们内心的先人之见，我们怎么能够说一个力比另一个力更加或更不“真实”呢？重要的是，我们能否通

过考虑力的效应而对物体如何运动作出一个好的描述，而且只要两种虚构力被包括进来，转动系中的物体运动就能够运用牛顿定律并不再理会参考系的运动而正确地计算出来。这对在实用工程学和物理学中使用这些虚构力是有效的鼓励。但是为什么在一个转动参考系里惯性就不同了呢？科里奥利力和离心力究竟是从哪里来的呢？贝克莱和马赫的回答实际上简单地就是“来自不变的恒星”。爱因斯坦则试图把答案建立在数学的基础上。

相对论背后的哲学

爱因斯坦是在学生时代知道马赫的书的，他后来给马赫写过信，热情地表达了自己的敬佩。马赫在 20 世纪的头 10 年里也对爱因斯坦的狭义相对论同样热情。在狭义相对论里惯性系有着与在牛顿力学里基本相同的地位。在牛顿看来，惯性是物体的固有性质；而在马赫看来，惯性是由“不变的恒星”造成的。按照马赫的思想，认为地球在相对于遥远的恒星转动，和认为恒星在围绕静止的地球转动，这两者并无不同。无论取哪种认识，假如恒星没有了，惯性将随之消失，而转动所产生的效应，如地球在赤道处的隆起，也都将消失。

马赫的论证与贝克莱相比几乎并无新意，但他生逢其时，是在 19 世纪 80 年代谈论运动、空间和不变恒星这些和贝克莱一样的话题，而那时正是对空间、时间和宇宙本质的深刻再思考即将到来之际。像贝克莱一样，马赫也认为向外的力，即离心力，是在物体相对于不变恒星转动时产生的，但他比贝克莱走得稍远，提出“是认为地球在绕自己的轴旋转，还是认为地球不动而那些不变的恒星在围绕它旋转，都无关紧要”。只要有相对转动就有向外的力，这种力就会使地球在赤道处隆起。爱因斯坦通过马赫的书了解到这些思想，也正是他给出

了“马赫原理”这个名称。它摈弃了牛顿的绝对空间概念，主张所有运动都是相对的，这个思想对狭义和广义相对论两者的建立都有重要影响。但是对它又一直有争议，予以谴责的人中就有列宁和伯特兰·罗素(Bertrand Russell)*。

爱因斯坦广义相对论的基石之一是等效原理。牛顿第二定律是力等于质量乘以加速度，其中的质量是表示物体反抗改变其运动的量，即惯性质量。两个物体之间的引力也依赖于质量，这个质量能越过空间而影响别的物体，称为引力质量。引力也与质量成正比，一个物体的引力质量和惯性质量是无法区分的。为什么这两个质量必定相等呢？毕竟一个是表示物体的“外延”即拉其他物体的能力，而另一个是表示其“内涵”即被其他物体推时的反抗，从逻辑上讲它们本该大不相同。爱因斯坦认为，引力质量与惯性质量相等，是因为惯性力本身实际上就起源于引力。

对此可作如下论证。远处的恒星以其引力之手保持着对地球上一切物体的控制。当我们试图改变一个物体的状态，即在直线上加速它或使它转动时，它是在那个引力网中加速，某种反作用就会产生我们在公园的滑道或旋转木马上体验过的那种反冲力。但那是何种反作用呢？一个“明显”的可能性是，干扰一个物体的运动会造成某种引力小波动，并向外蔓延到远处的恒星和星系，于是它们就作出某种回应，试图维持该物体的原状。但是，如果像爱因斯坦方程所要求的，信号不可能传递得比光速更快，那又怎么能像我拨一下桌上的铅笔，而铅笔即刻就知道它该作多大反抗，或者说有多大惯性呢？答案来自于这样的认识，物质、空间和时间连成了一个不可分割的整体。按照

---

* 英国著名哲学家、数学家和逻辑学家。——译者

对爱因斯坦工作的这种马赫式解释，我此时所在之处的时空结构取决于宇宙中所有物质的平均分布。而此时此处一个物体的惯性，即它对一个作用力的反抗，则取决于该时空结构。原理——如果你喜欢的话或叫哲学——似乎已足够清楚。麻烦是在爱因斯坦试图把这个哲学输入广义相对论方程时出现的。

当爱因斯坦着手建立广义相对论时，他本想把马赫原理包括进去，或者说，他本想马赫原理成为他的理论的自然结论。他以为自己已经成功了。广义相对论方程应该能包括远处质量(我们现在是指遥远的星系而不是不变的恒星)的影响，而加速度、惯性力和转动是相对于这些遥远天体而言的。但是，怪事之一是，爱因斯坦方程只有对一种特殊类型的宇宙才能给出符合马赫原理的结果，这种宇宙中有足够的物质使空间弯回到自己上面，因而宇宙是闭合的。如果这是对的，那就意味着我们是生活在一个极大的黑洞里面，而且这对现在关于宇宙起源和结局的争论也有重要意义，这种争论将在第三篇里介绍。马赫从来没有喜欢过广义相对论，1916 年去世前还在准备写一本书来反驳这个理论，自他死后关于相对论是否确实以令人满意的数学方式包含了马赫原理的争论时起时伏。有关的方程式复杂得可怕(这是马赫反对的原因之一)，而且一直到最近为止争论只限于在数学家和哲学家中进行。但是现在看来已经得到了那些方程的一个解，这似乎是在为爱因斯坦辩护，而且已经有可能设计出一个可行的实验，来检验被结合到爱因斯坦理论中的那个马赫原理翻版的精度。

如果作为静止标准的局域参考系果真是由宇宙中所有物质的某种平均效应决定的，那就应该有某种方式来安排物质的局部集中，以在静止标准中产生出可测量的变化。设想有一个物体放在一个很大的物质球壳里，球壳本身在相对于遥远星系旋转。如果马赫原理正确，就

应该有一种来自旋转球壳的小拖曳效应，试图拖着它里面的物体一起转。而任何宣称要包含马赫原理的理论，比如广义相对论，就必须首先能预言这种球壳内对惯性系的拖曳，然后给出这种效应应该有多大。早在1918年，蒂林(H. Thirring)就解出了有关的相对论方程，但只是在所谓的“弱场近似”条件下，也就是说仅能应用于相对较弱的引力场。结果表明，爱因斯坦方程的确能给出在转动的物质球壳内、或在像地球这样的转动物体附近的马赫式惯性系拖曳效应，但是所预言的效应强度太小，无法测量。

20世纪50年代，普林斯顿大学的迪克(Robert Dicke)重新挑起了争论。他指出，没有人解过对应于强引力场的方程，强场就要求有非常大质量的球壳，那么按照马赫的思想，球壳就会支配它里面的惯性系。在过去20年里，耶鲁大学的科恩(Jeffrey Cohen)和许多同行一起步步为营地向强场难题进攻。他们已经证明相应的相对论方程是有解的，并且的确作出了关于大质量转动球壳内转动物体行为的“马赫式”预言。就在这艰难的数学跋涉的同时，关于宇宙的新发现已经提供了检验由方程中得出的定量结果的实际可能性。

遗憾的是详细情况太复杂了。共有10个联立的非线性微分方程，听上去就知道很难对付。爱因斯坦建立了方程却没能解出它们。他知道，如果这些方程没有解存在，广义相对论就不能解释摆的行为或桶里水的涡动，于是就会被抛弃。有趣的倒是，爱因斯坦在世时就已经存在现在用来解决其未竟其功的问题的数学方法。方法是法国数学家嘉当(Elie Cartan)发明的，他从1869年活到1951年，但他的工作直到他快去世时才广为人知。他关于微积分和微分几何的工作，现在被认为是20世纪最有创造性的数学工作的一部分，为科恩及其同行提供了所需要的数学工具。科恩是生活在这个计算机主宰的年代，

他却自豪地指出，他的所有计算都没有使用比铅笔和纸更高级的东西，当然要相当多的纸。科恩的解，假定能经得住其他数学家的严格审查的话，就填补了广义相对论中一块令人窘迫的空白。这个解还为数学物理学家提供了手段，去对付高密度旋转天体的内部过程的难题，表现为脉冲星的中子星就是这种天体。这个解还提供了理解你杯中咖啡的框架——咖啡怎么知道它相对于宇宙中所有物质的平均分布而言是在哪里，是在如何运动。你下次搅咖啡时不妨想想。

涉及强引力场的转动也出现在自转中子星即脉冲星和黑洞的场合。作为一种可能性，现在正从爱因斯坦将近 70 年前建立而未解出的方程中得出的数据，将给出关于这些天体观测性质的预言，从而对理论作出检验。另一方面，由于蒂林的时代以后技术的改进，即使是弱场解也可以付诸检验。斯坦福大学的一个小组自 20 世纪 60 年代以来一直在筹划一个环绕地球轨道上的失重陀螺仪的实验。地球自转对惯性系的拖曳作用将影响这个装置。在失重的情况下，这些细微的影响有可能被检测出来，因为没有摩擦力来制造混乱。当然，即使斯坦福陀螺仪被送上了轨道，它能否测出蒂林预言的效应仍难断言。但仅是这种可能性就已经显示了过去几十年所取得的巨大进步。

不过，我们大概并不真需要这些精巧细微的东西。我在前面提到的马赫原理的宇宙学含义是什么？马赫像贝克莱一样，谈的是“不变的恒星”，因为那时没有人知道我们银河系只是散布在空间的千千万万个星系之一。但即使在哈勃的星云巡天表明银河系只是宇宙中的一个岛之前，观测已经显示银河系是一个扁平的盘状系统。赖特已经有此认识，而赫歇尔等人在马赫出生以前就已为这一认识建立了相当可靠的基础。我们银河系是一个扁平盘状系统，它的形状是由转动和离心力决定的。假如马赫或另一位 19 世纪的天文学家采用关于银河系

形状的证据，并按照马赫原理的思路来考虑这些证据，就应该可以作出如下论证。由相对于不变恒星的运动所造成的离心力使桶里的水沿桶内壁上升。马赫的假设是，这些力的出现是因为水在相对于不变的恒星转动。而现在知道那些“不变的恒星”本身是一个被离心力扭曲的转动系统的组成部分。要么牛顿是对的，不变恒星的系统作为一个整体在相对于绝对空间转动；要么马赫是对的，那就必定有某种远在银河系之外的物质分布来建立一个参考系，银河系的转动正是相对于那个参考系而言的。

在 19 世纪，依据关于银河系的已知事实和马赫原理，本可预言存在着数量巨大的遥远星系。假如这种预言曾经有过，哈勃在 20 世纪 20 年代建立宇宙尺度的先驱性工作，恐怕就不会那么惊人，而会被看作是对马赫原理和广义相对论两者正确性的证实。

真是挺好，可是从来也没有人作过这种预言，的确没有。而且我在想，究竟有多认真的预言能产生于另一个哲学思辨的光亮(或许我应该说是黑暗)中——这个思辨存在于 19 世纪以及更早年代，事后看来，它应该导致关于宇宙正处在演化之中的预言。

黑暗的夜空

我们所生活的宇宙不仅黑暗和空虚，而且寒冷。寒冷是黑暗的另一种说法，因为热的物体如太阳，会以光的形式辐射能量。自然界的基本定律之一即热力学第二定律说，能量只能从较热的物体转移到较冷的物体，所以经过足够长的时间后两物体的温度会变得一样。这样看来，在寒冷、黑暗的宇宙里点缀着明亮、高温的物体(即恒星)这一事实就非常令人吃惊。这至少也应该是惊人的，但是我们太习惯于那黑暗的夜空，就那么默认了这一现实，而从不停下来想想宇宙是处在

一种很不正常的状态，按照最基本的物理定律这种状态是不可能永久，甚至不可能延续很长时间的。夜空的黑暗是一个需要解释的难题，而对这个难题的解释就揭示出我们这个宇宙的很基本的东西。

这个难题常被称作“奥伯斯佯谬”，其名称是在20世纪50年代由宇宙学家邦迪(Hermann Bondi)普及的，用以纪念 的人物奥伯斯(Heinrich Olbers)是一位德国天文学家，从1758年活到1840年，在1823年写了一篇讨论该难题的标志性论文。遗憾的是这个名称并不恰当。正如马赫不是第一个讨论“马赫原理”的人一样，奥伯斯也不是第一个为“奥伯斯佯谬”伤脑筋的人；更糟的是，对于那些在此类事情上喜欢高度精确的人来说，这个难题其实根本就不是一个佯谬，而就是一个难题，一个特别有趣和内涵丰富的难题。但是，名称已经定了。那么，什么是奥伯斯佯谬呢?

这个难题是基于三个假设：宇宙是无限大的；宇宙中到处分布着与太阳和银河系里的恒星相差不多的恒星；宇宙是永恒不变的。在这种情况下，我们往天空中看时就应该在任何方向上都能看到恒星，那么我们为什么会看见恒星之间的黑暗空间呢? 如果你想象在任何一个方向上画一条直线往外伸进宇宙，也就是一条“视线”，那么在一个均匀地分布着恒星的无限宇宙里这条线总会在什么地方碰上一颗恒星。由于光沿直线行进，该恒星的光就应该沿视线来到我们这里。不管往哪里看，我们都应该看到一颗恒星，整个天空就应该是光辉普照。

这只是一个简单的、被天文学家归为所谓“挥手”型的论证。但是用一点初等几何就可以把“佯谬”放到可靠的数学立足点上。想象地球是在一个包含有许多恒星的大球的中心。在这个球的边界上也分布有恒星，为论证起见，我们可以想象球被一层薄壳包着，就像橘子

上的皮，而壳上有着一定数量的恒星。如果星的数量足够大(这在无限大的宇宙里是容易做到的)，我们就可以说，每颗星都对从地球上看的夜空贡献大致相同的平均亮度。每颗恒星的亮度当然是正比于其真实的绝对亮度，而反比于与我们距离的平方。所以整个恒星壳即整个皮层所贡献的亮度就等于壳上恒星的数量乘以标准亮度再除以距离的平方，而球壳上每颗恒星的距离就是球的半径。

有着相同真实亮度的恒星，离得较远的看上去就显得较暗。所以你可能以为这种缀有恒星的球壳也是半径越大就显得越暗。但是按照几何知识并非如此。越大的壳固然是离我们越远，但正因为它大，它上面的恒星也越多。如果恒星是在无限大的空间里均匀地分布着，就很容易算出每层壳上有多少恒星。由于球面的面积是与半径的平方成正比，每层壳上的恒星数目就也与半径的平方成正比。每颗星的亮度是随半径平方而减弱，但每层壳上的恒星数目的增多正好抵消了这种减弱。[2] 结果是每一层球壳都给夜空贡献同样的亮度！

这可就比那个挥手型论证所给出的要糟多了，既然是无限大的宇宙，就必定有无限多个此种壳层，它们的合效果就将是一个无限亮的天空。当然，较近的恒星会遮挡较远恒星的光，所以天空不会无限亮，但总应该处处都有一颗典型恒星的表面那样亮，也就是像太阳表面那样亮。难题已经不仅是天空为什么在夜晚是黑暗的，而是为什么在白天也这么暗。

有一种论证方法叫做归谬法，就是从某个基本假定出发推演出一个显然是荒谬的结论，从而证明最初的假定是错的。奥伯斯佯谬就是由简单的假定不可避免地导致天空必定明亮的结论，而事实上天空却是黑暗的。所以其基本假定中至少有一个必定错了。你不可能这样辩解，说银河系只是空中一个岛，所以对恒星的计数只能到银河系边界

为止。只要把“恒星”一词换成“星系”，则前面的论证依然成立。奥伯斯佯谬所揭示的是，宇宙并不处于热平衡状态，尽管所有星系中的所有恒星都在努力倾泻能量给它加温，它的绝大部分仍是寒冷的。

你对发现夜空黑暗的惊讶程度依赖于你对导致那个荒谬结论的最初假定的相信程度，而这又非常强烈地依赖于你成长时所接受的文化和所养成的科学思维习惯。的确，首先注意到夜黑难题的那些天文学家很幸运地不知道问题的严重性，所以没有表示出什么惊讶。这也许就是他们的名字没有被用来给“佯谬”命名的缘故。马萨诸塞大学的天文学家哈里森(Edward Harrison)已经作了最大努力来整理这方面的历史记录。[3] 这肯定有助于了解那个起错了名的奥伯斯佯谬究竟对宇宙作了什么揭示，或者说它的基本假设中究竟是哪个错了。

谜的由来

第一个基本假设说的是宇宙的外延无限，这是一个相对较新的概念。古代人乐于相信平坦的大地搁在一只大龟背上之类的思想，从不去想无限宇宙的可能性。后来公认的图像是地球被一系列载有行星的天球环绕着，可见的恒星则全都固着在最外一层天球上，就像那很大但肯定是有限的教堂穹顶的微光。英国人迪格斯(Thomas Digges)在1576年写的文章中接受了托勒玫的这一思想，但把那固定恒星的单一水晶球分散成了无限延展的空间。迪格斯应该享有把无限的概念引入近代宇宙图景的荣誉[虽然公正地说来希腊思想家德谟克利特(Democritus)也考虑过无限空间]。迪格斯还意识到需要解释为什么在一个无限大的宇宙里夜晚的天空是黑的。如哈里森所报告的，迪格斯相信，夜空黑暗只是因为遥远的恒星太暗弱而根本看不见。这在16世纪后期是一个完全合意的“解释”。我们现在知道它站不住脚，

因为一个球壳上恒星的光应该加起来。但是迪格斯引进了充满恒星的无限宇宙的思想，并且注意到需要解释为什么会看到恒星之间的黑暗天空。他把天文学家送上了通往奥伯斯佯谬的道路。

开普勒于1610年接过了这道难题。根据哈里森所搜集的历史记录，开普勒第一个意识到夜空黑暗与布满明亮恒星的无限宇宙的思想直接冲突。也许本该是“开普勒佯谬”，但是开普勒从来没有在这种意义上提出难题，因为他相信宇宙必定有一个边界，于是他认为夜空的黑暗正是对宇宙外延有限的证实。开普勒实际上是说，当我们的目光穿过恒星之间的天空时，我们所看到的是围绕宇宙的黑暗墙壁。但当难题于1721年传到以彗星成名的哈雷手上时，他又返回到那个恒星太遥远以至于暗得看不见的思想。于是舞台就让给了另一个人来首先把难题表述成与此处所述相差不多的形式。这个人不是奥伯斯，而是一位瑞士天文学家德切塞克斯(Jean-Philippe Loÿs de Chéseaux)，他出生于1718年。

德切塞克斯是一名神童，受教于本身也是一位杰出学者的祖父，18岁时就负责建造瑞士的第一座天文台，在18世纪40年代初就因发现两颗彗星并写了一本有关彗星观测的书而成名。正是在那本书的附录里他给出了自己对夜黑难题的表述。他没有作那种挥手论证，而是尽其所能地通过把恒星的亮度与太阳和行星的亮度作比较来计算恒星的距离，并且计算了一个像太阳那样的恒星在适当的距离上将覆盖天空的多大一块。他一步一步地到达了这样的结论，如果恒星在宇宙中分布的疏密程度与我们附近恒星的分布状况相似，那么在一个大到(用现代单位)10^{15}光年的球内，我们就会在天空的任何一个方向上都看到恒星。所以，要么恒星不是在无限的空间均匀分布，要么就是有什么东西在阻挡我们看到遥远恒星的光。[4] 与哈雷和迪格斯不同，德

切塞克斯认识到，几何学保证了遥远恒星的暗弱恰被它们的巨大数量所抵消，所以不能简单地说遥远恒星会太暗弱而看不见。他提出，真空会吸收恒星光的能量，所以光在越过空间行进时会变得越来越暗弱。这正是奥伯斯在他于 1823 年讨论夜黑难题时所采纳的观点。

奥伯斯于 1758 年出生于不来梅附近，在格丁根大学学医，也听物理学和天文学课。他后来一直当医生，也一直保持着对天文学的兴趣。他把自己家的二楼变成了一个天文台，用自己的望远镜发现彗星和小行星。在 64 岁从医生职业退休的那一年，他发表了对夜黑难题的讨论。他的设想是在恒星之间有一种冷物质，即一种弥漫于空间的稀薄的气体—尘埃糊，于是难题就可以解决，因为这种物质会吸收遥远恒星的辐射，从而遮挡了它们的光和热。这种论证在一定范围内讲得通，但是显然不够。那些物质糊吸收了能量会怎么样呢？当然会变热，最后就会热到也辐射能量。在一个无限和不变的宇宙里，应该已有足够的时间，其实是无限长的时间，来以这种方式使那些物质糊加热到与恒星的辐射处于完全平衡。在这种状态，那些物质糊所吸收的能量与所辐射的能量精确相等，于是又返回到了为什么夜晚天空黑暗的难题。奥伯斯显然没有充分考虑宇宙不变这一假定的含义。他的文化条件使得这个假定在他脑中根深蒂固，他大概从未有过疑问。但是一旦看清了那些含义，就得开始考虑恐怕正是必须抛弃这个假定，才能避免奥伯斯佯谬的荒谬。

解开难题

借助于能够探测深层空间的望远镜，哈勃和他的后继者已经揭示出，我们所生活的宇宙至少也大到在夜黑难题的意义上可被看作无限大，而且宇宙中物质的分布也非常均匀，在所有方向上尽望远镜能力

所及都能看到星系和星系团。于是只剩下一个不变性假设可被怀疑，而哈勃当然也表明了宇宙并不是恒定的。宇宙是在膨胀，星系在随着时间的流逝而相互分离得越来越远。来自遥远星系的光被红移了，波长被拉长，相应的能量减小。这能够解释所有那些遥远恒星和星系的能量去向从而回答夜空为什么是黑的吗？也许能，但恐怕至多也只是答案的一部分。其余部分则涉及一个更突然的改变，那反倒是 17 世纪和 18 世纪的哲学家和天文学家应该比我们更宜于接受的。假定宇宙有一个确定的开端，就是说它有一个“边界”，不是在空间上而是在时间上。

由于光的速度是有限的，我们看遥远的恒星和星系时所看到的就是它们很久以前的样子。按照我们地球上的钟，距离为 5 光年的恒星发的光要用 5 年时间到达我们这里，而距离为 50 000 光年的星系的光就得要 50 000 年。于是“佯谬”就只是在这样一种情况下才成其为难题，那就是宇宙中的恒星和星系已经有足够长的时间把足够多的光送到我们这里，使得夜晚的天空明亮。但是如果有一个确定的创世时刻，而那个时刻距现在又不那么久，那么就没有什么佯谬了。这就是问题的关键所在。

事后看来，令人惊讶的是牛顿和他的同时代人没有猜到对难题的这种解释。光速的有限性已在 1676 年由丹麦天文学家勒默尔（Ole Roemer）通过研究木星对其卫星的掩食而确定，并且已为牛顿所知，因为后者在他 1704 年出版的《光学》一书中提到过。当哈雷在 1721 年的皇家学会会议上报告他关于夜黑难题的两篇论文时，牛顿是会议主席。牛顿没能指出，此后两个世纪里也没人能指出，只要假定相当久以前没有恒星存在，难题就可解决。

使这种情况显得更令人费解的是，那时教堂里正讲授着创世发生

于公元前4004年。与牛顿同时的任何一个真正相信这种说法的天文学家马上就可以说，凡是距离比4004加1721光年更远的恒星在地球上都还看不到，而一个半径只有5000多光年的空间远不能容纳足够多的恒星来把夜空照亮。哲学家没能这样指出来，也许说明他们对“正统”的创世日期是多么缺乏信念，倒不是他们忽略了其含义。但是爱伦·坡(Edgar Allan Poe)曾在1848年指出，向空中远眺也就是在时间上回顾，就可以看到恒星诞生以前的黑暗状态，这个话题在1907年被爱尔兰学者达尔伯(Fournier d'Albe)接过去详加发挥。[5]

所以，奥伯斯佯谬有两个解，或者说至少有两种对于宇宙变化的认识。一是遥远星系光的能量已经被由宇宙膨胀造成的红移减弱到不能照亮夜空；二是宇宙存在的时间还短得不足以达到热平衡态，也就是说还没有足够的时间来让恒星的辐射充满它们之间寒冷黑暗的空间。400年前迪格斯年代的那场争论，现在仍未结束，因为上述效应中哪一种更重要的问题仍在讨论之中。

一派观点是由邦迪强烈主张的，正是他在20世纪50年代使人们对奥伯斯佯谬的兴趣复活。那时的邦迪又是稳恒态宇宙模型的创造者之一。这个颇引人好奇的模型大意是，虽然宇宙在膨胀，星系团在互相分离，但宇宙仍然可以实质上保持不变，因为新物质可以从虚空中产生出来并形成新的恒星和星系，以填补老星系移开后出现的空缺。这个模型在20世纪50年代得到了一些支持，但到60年代就败下阵来，因为有证据表明确有一个创世的时刻，宇宙是诞生于大约150亿年前的一次大爆炸。显然，在一个名副其实的稳恒态宇宙里，时间和空间都是无限的，奥伯斯佯谬才真是一个问题，唯一的解决办法就是星光的能量随红移而消失。

最近，邦迪和我一样倾向于从热平衡的角度来讨论夜黑难题。[6]

恒星表面的温度大约是5000开，而星际空间的温度只有区区几开，即绝对零度(－273摄氏度)以上几度。任何一个恒星或星系都不可能永久存在。恒星的光和热来自把氢转变为较重元素的热核反应。这种反应所能提供的能量是限定的，而恒星的寿命只有几十亿年。显然恒星并没有释放出足够的能量来把整个宇宙加热到使夜空明亮的程度。看来我们的思路应该完全变换。恒星是在那里不停地猛烈燃烧，当到达热平衡态时(如果真能到达)，其温度就不会是5000开而是绝对零度以上几度。难题现在掉了个头，不是“为什么星星之间如此黑暗？”而是“为什么竟还能有明亮的星？”而答案再次是，宇宙是产生于一个与热平衡态非常不同的状态，而且它所产生的时刻距今不会超过几代恒星寿命。

哈里森说得更有诗意：“穿过群星之间的空隙，我们回顾宇宙的开端。”我们在那些空隙里究竟看到了什么呢？一种非常微弱的辐射噪声，一种非常均匀的能量分布，等效温度是约3开。这种辐射是创世的回声，是比任何恒星都更致密的火球的能量经过极高红移后的残余。夜黑的成因终于找到了，这就是宇宙的膨胀。

哲学家积许多世纪之沉思，不能预见这种背景辐射的存在，不能说明创世瞬间的本质。但是以夜空黑暗为据，他们有可能也应该能意识到，宇宙会有一个起始，而且与恒星的年龄相比不会早得很多。他们一再坐失良机，既没有运用马赫原理来推断宇宙比他们所想象的要大得多，又没有透过奥伯斯佯谬来领悟宇宙比他们想象的要年轻得多。哲学家解释宇宙的记录的确远不敢恭维，然而科学家对宇宙的思考也常常如此，即使是在20世纪。似乎有着一种普遍的态度，一种当代的文化制约，就是总觉得那些抽象的理论固然引人入胜，却不能真正向我们揭示这个宇宙的产生和演化。从20世纪20年代到60年

代，正是这种态度使形而上学家，同时也有意无意地使早期宇宙学家，没有能抓住自己工作的形而上学意义。所以我们也许不该苛责前辈哲学家没有充分利用他们已经得到的证据。然而惊人的事情的确是，少量简单的观测和一些仔细的思索，就足以深刻揭示我们这个宇宙的实质。但是时代潮流已经把形而上学家再次放到一边，并重新回到 20 世纪 20 年代和 30 年代天文学家的工作，是他们把爱因斯坦的广义相对论和哈勃对膨胀宇宙的观测结合起来，促成了现代宇宙学的诞生，并开始探索大爆炸本身的奥秘。

注　释：

1. 此段贝克莱的话和上面一段牛顿《原理》中的话均引自席阿玛的书《宇宙的统一》(*The Unity of the Universe*)。

2. 稍想一下就知道这是为什么，每颗在一定距离上的恒星的亮度与距离平方成反比，是因为该恒星的光在不断地充满一个在空中胀大的泡，而泡的表面积与半径平方成正比，几何是对称的，所以距离被消掉了。

3. 先后几次发表，最近的刊于 *Science*, Vol. 226, pp. 941—945，其中有关于奥伯斯佯谬的所有主要参考文献。

4. 当然可以再次把“恒星”换成“星系”而作出等价的论证。

5. 详细情况已由哈里森给出。

6. 见梅塞尔(H. Messel)编《今日科学》(*Science Update*)一书中邦迪的文章，该书由牛津 Pergamon 出版社 1983 年出版。

第二篇

大 爆 炸

每当一个年轻人告诉我他想搞宇宙学时，我总是很惊讶，因为我认为宇宙学是可遇而不可求的。

——麦克雷


PDG

第四章　宇 宙 蛋

现代宇宙学开始于爱因斯坦的广义相对论和他 1917 年的第一篇宇宙学论文，他在文中迫使相对论方程去适合当时的宇宙静止而永恒的错误信念。但是文中的思想并非凭空而来。种子在 19 世纪中期就以有关弯曲空间和非欧几何的概念的形式播下。这些概念可能与真实宇宙有某种联系的想法甚至也在该世纪末由英国人克利福德(William Clifford)表达过，他不幸只活到 34 岁就于 1879 年(正是爱因斯坦出生的那一年)去世，没来得及把自己的思想发展到构造出第一个现代宇宙学模型。

非欧几何研究的先锋还有高斯(Karl Friedrich Gauss)。这位才华横溢的德国数学家生于 1777 年，卒于 1855 年。尽管他对其中平行线的行为将与日常常识和欧几里得几何里不同的新几何的研究相当有限，“非欧几何”这个名称却是他首创的。对建立完整的非欧几何贡献最大的，是俄国人罗巴切夫斯基[Nikolai Ivanovich Lobachevski，他的名字在莱勒(Tom Lehrer)一首有趣的小曲里流芳百世]和匈牙利

人鲍耶(Janos Bolyai)，两人在19世纪20年代相互独立地工作，并各自得出一种特殊形式的非欧几何。1854年，另一个德国人黎曼(Bernhard Riemann)检查了几何学的整个基础，确定了各种可能的互不相同的几何的整个范围，其中的每一种都和其他任何一种一样有效，而欧几里得几何只是大量可能性中的一个例子，于是非欧几何作为一个数学分支就有了坚实可靠的立足点。黎曼工作的普遍性还在于考虑了把几何学推广到三维以上领域的可能性。随着黎曼的工作而发展起来的新数学成了爱因斯坦至关重要的工具。正是黎曼几何和别的工具一起，使得从数学上研究这样一种理论宇宙模型成为可能，这是一个四维的球即所谓“超球”，其中我们能直接体验的三个空间维度均以恒定的曲率穿过第四个维度而弯曲，并且弯回到与自己相接，正像在一个球面上画的“直线”会弯回到与自己相接那样，这样就成了一个“闭合”的宇宙。

克利福德把这些思想介绍给了英语世界，他把黎曼的工作译成了英文，并且一定也像黎曼一样意识到“有限而无界”的宇宙即球面的高维对应物之可能性。他在1870年向剑桥哲学学会提交了一篇论文，其中谈到“空间曲率的变化”并作了这样的类比：“空间的一些小部分的性质类似于一个平均而言是平直的表面上的小丘，通常的几何定律对那些小部分是不适用的。”[1]诺思(J. D. North)在他的书《宇宙的量度》[2]中提到了到1915年为止的半个世纪中发表的80多篇关于非欧几何的静力学、运动学和动力学的学术论文。那些作者全都有可能来一个思想飞跃，提出可以把自己的方程式应用于真实宇宙。但是他们全都没有做到。那个思想飞跃不得不等待一个人来实现，他从不畏惧进入未知的科学领域，他借助于黎曼及其同代人提供的工具构造出了一个全新的引力理论，也构造出了宇宙。爱因斯坦使球开始滚

动。但是还要等上 10 年人们才开始认真对待用那些方程式描述真实宇宙的主张，还要等上半个世纪宇宙学家们才认识到那种描述是多么“真实”。

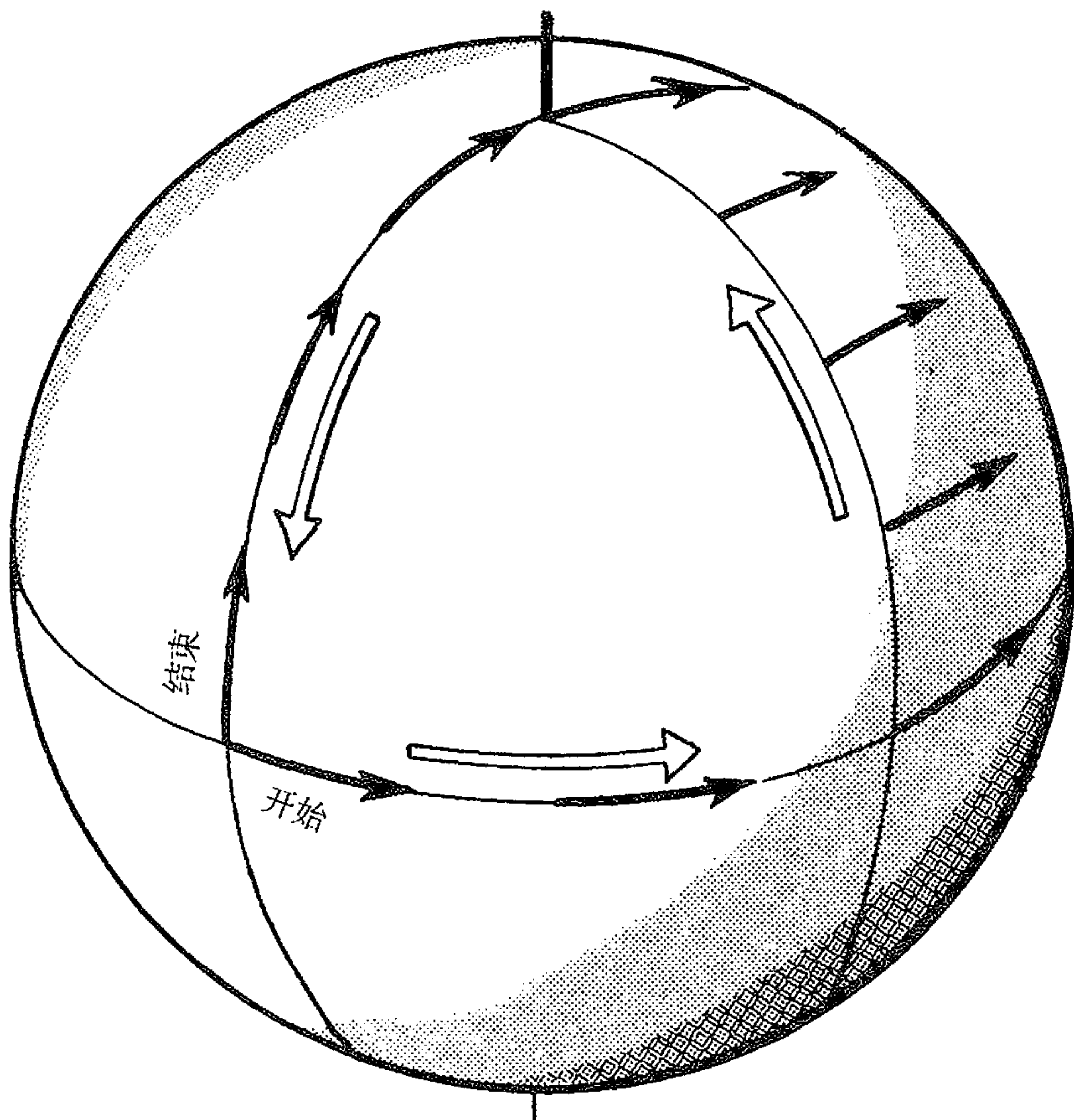


图 4.1 非欧几何奇特性的一个简单例子。在一个球面上有可能让一个小箭头（一个矢量）先绕着赤道旅行，再到达极点，并返回到出发点。在行程中的每一步都非常小心地使箭头保持指着同一个方向。但当箭头回到出发点时，它的指向竟是正北而不是正东！

宇宙学诸父

也许最好把爱因斯坦看作现代宇宙学的“祖父”而不是“父亲”。他指出了大方向并使球开始滚动，但却是其他人来使球加速，并且是沿着有违他原意的途径。从 1917 年起，所有关于宇宙的新数学模型都包含膨胀，甚至在哈勃和赫马森不容置疑地确定真实宇宙的膨胀之前就已是如此。

这些新模型的第一个紧接着爱因斯坦的静态宇宙模型而来，提出者是德西特，那位把广义相对论转达给在伦敦的爱丁顿的荷兰人。1915年关于爱因斯坦新理论的消息到达莱顿时，德西特已经是一位经验丰富的老资格天文学家。他生于1872年，在格罗宁根大学和开普敦皇家天文台学习，1901年获博士学位，1908年成为莱顿大学的理论天文学教授，后来还当了莱顿天文台台长，1934年死于肺炎。德西特教授是考虑狭义相对论在自己专业中的应用的很少几个天文学家之一，该理论在发表后的头10年里被普遍认为只有数学家才会有兴趣，而谈不上什么实际作用。他也应该有可能是爱因斯坦之后第一个应用广义相对论的人，这当然部分地由于他是有幸最早得知这一新理论的人之一。

当爱因斯坦寻求一种依据广义相对论的宇宙描述时，他的确找到了一种描述，即他的方程的一个解。他的含有宇宙常数的静态模型似乎满足了需要。但是德西特在1917年寄给伦敦皇家天文学会的一篇论文(该文曾由当时的学会秘书爱丁顿以极大的兴趣宣读)里证明，方程还有另外一个解，一个描述不同的模型宇宙的解。很明显，这两个解都不能代表真实宇宙。我们现在知道这一点，而这也不被看作是什么问题。但在当时，这却像是对爱因斯坦理论的一个打击，因为如果这理论竟允许有不同的却又都符合基本方程的宇宙，那么它恐怕就不能告诉我们有关真实宇宙的多少情况，如果它还能告诉一点的话。宇宙的膨胀一被发现，这种看法就不能成立了。天文学家也意识到，本可在哈勃宣布红移—距离关系之前至少10年就作出关于宇宙膨胀的预言。

德西特的宇宙像爱因斯坦的一样，在数学意义上也是静态的(还像爱因斯坦的一样包含宇宙常数)。但与爱因斯坦宇宙不同的是，它

里面根本没有物质，它是一个空空如也的宇宙的数学描述。很难明白在一个完全真空的宇宙里"静态"是什么意思，因为没有任何东西能用作标志来测量运动。当理论家尝试给德西特宇宙洒进几滴物质时怪事出现了——那些物质点滴即检验粒子立即四散而去。还有，当他们计算一个检验粒子发出的光在另一个检验粒子看来将会怎样时，竟发现了与粒子间距离成正比的红移。德西特宇宙看上去是静态的只是因为它是真空的，对一个含有少量物质即在空中这里和那里散布着少数星系的宇宙，天文学家本应能精确预见哈勃和赫马森要到20世纪20年代末才能找到的红移—距离关系。很久以后，爱丁顿这样来概括最早的两个相对论宇宙模型的区别：爱因斯坦的宇宙有物质却没有运动，德西特的宇宙有运动却没有物质。[3]

爱丁顿是当时认真看待含有微量物质的德西特宇宙的膨胀这件事的很少几个人之一，根据20世纪20年代初主要由斯里弗得到的很有限的红移资料，爱丁顿推断德西特模型确实对真实宇宙有所揭示。他后来也拿出了自己的一个模型，其中的宇宙先处于静态很长时间（也许是无限长时间），就像爱因斯坦宇宙；然后随着星系的形成而开始膨胀，又像德西特宇宙。但这个模型很快就被看出与真实宇宙的性质毫不相干。一旦红移—距离关系开始显得是真实宇宙的重要特征，爱因斯坦和德西特对自己的宇宙模型就作了重要的再度思考。1932年他俩联手共建了另一个宇宙模型，即爱因斯坦—德西特模型（不要与他们原来各自的独奏混淆），在一定意义上返回了根基。宇宙常数当初是为静止模型引入的，但是真实宇宙是在膨胀，所以去掉那个常数。以前的模型采用了弯曲空间（在德西特模型里时间也弯曲），但是没有直接证据表明空间是弯曲的，所以去掉弯曲空间（但不是去掉弯曲时空）。爱因斯坦—德西特宇宙是用广义相对论基本方程所能构造

出的最简单的宇宙。它如方程所要求的是在膨胀，膨胀着的空间是平直的，是狭义相对论的空间。由于模型中没有加进什么东西来阻止我们逆着时间回溯，模型就要求在很久以前有一个确定的创世事件，那时宇宙从一个数学点即从一个无限大密度的状态诞生出来，那种状态叫做奇点。

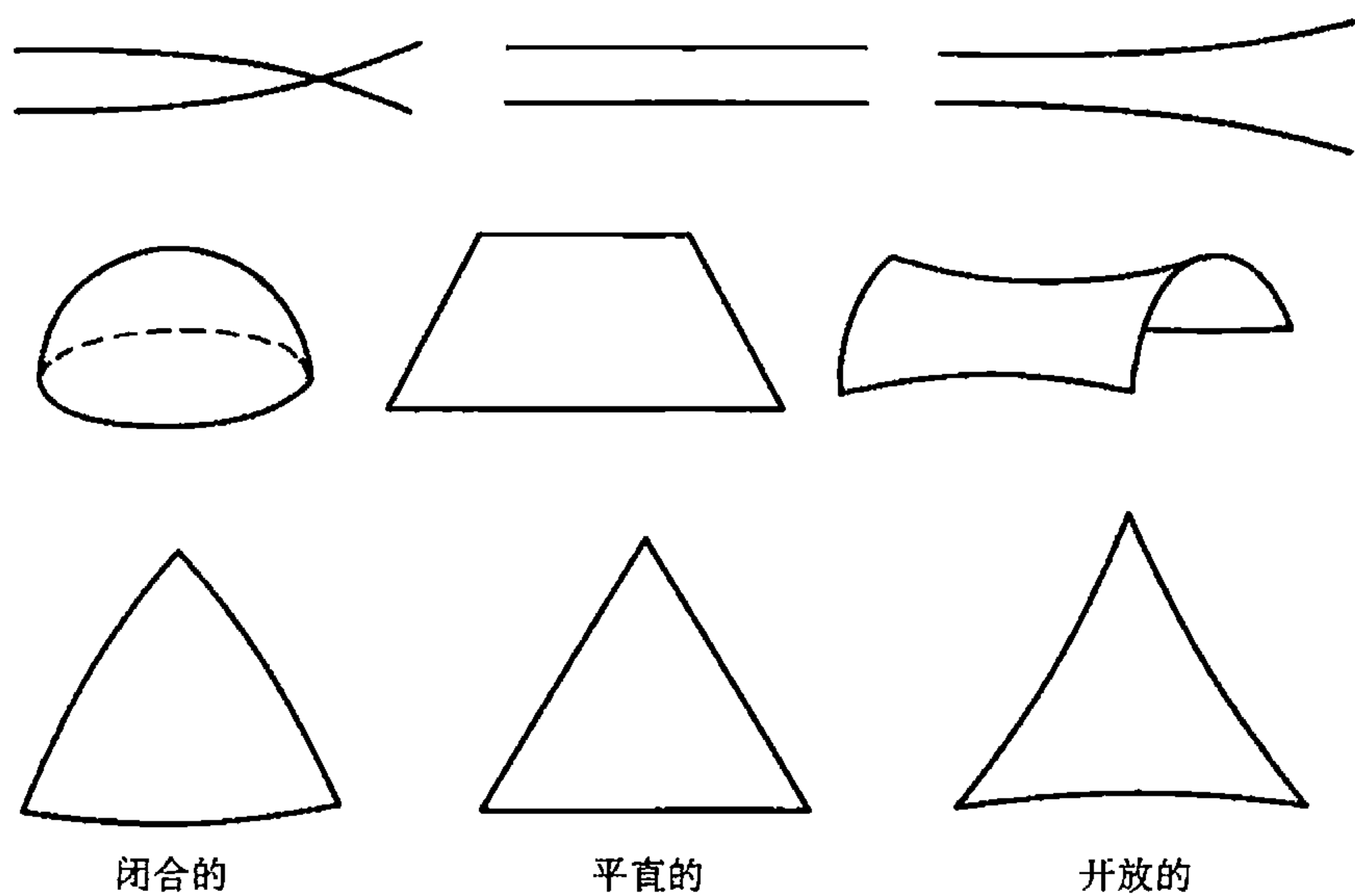


图 4.2　空间可以与三种基本几何的某一种相符。真实空间是三维的，但我们可以通过二维“宇宙”的曲率来看出三种可能性。

如果空间有着正曲率，即像球面那样是闭合的，那么“平行线”最终必定相交，三角形的三个内角之和大于 180°。

如果它的曲率为负值，即像马鞍面那样是开放的，那么平行线会越来越散开，三角形内角和小于 180°。

平直空间是一种特别情况，把闭合宇宙与开放宇宙分开。只有在这种很特别的宇宙里平行线之间的距离保持不变，三角形的内角和正好等于 180°。我们的宇宙很接近于是平直的。解释为何如此是宇宙学的一个基本问题。

作为方程的最简单的解，爱因斯坦—德西特模型是一个在讲课时很有用的例证。今天的许多宇宙学课程很合理地把它摆在首要位置，因为学生应该从最简单的例子开始，再逐步深入到复杂和有趣的情况。但是爱因斯坦和德西特对这个创作的自我感觉如何就很不清楚了。钱德拉塞卡(Chandrasekhar)在《爱丁顿》一书中记述了 1932 年

爱丁顿在爱因斯坦—德西特论文刚刚发表后所作的一番评论[4]：

> 一会儿后爱因斯坦过来和我待在一起，我对他批评了这篇文章。他回答说："我自己不觉得这篇文章很重要，但是德西特热中于它。"就在爱因斯坦走后，我收到德西特的信说要来见我，他还写道："你应该已经看到爱因斯坦和我的文章，我自己不认为那结果有多重要，但看来爱因斯坦是那么认为的。"

事后看来，爱因斯坦—德西特模型的关键特点，也是它被如此广泛地讲授的理由，就是它包含了那个创世瞬间，即我们现在所称的大爆炸。但是这两位先驱提出模型时的那种踌躇态度也许反映了他们自己对那个思想的不安，那个思想并不是他们的，而是独立地来自 20 世纪 20 年代的另外两位现代宇宙学奠基人，并且只在哈勃的红移—距离关系发表之后才受到注意。大爆炸得到充分重视是在 20 世纪 40 年代，直到 60 年代才成为宇宙学中的主导理论。

弗里德曼(Alexander Friedman)没能活到看见自己的工作被承认为对 20 世纪宇宙学的关键贡献之一。爱因斯坦强迫自己的方程去描述一个静止宇宙；德西特从方程中得出了一个静态但却是真空的宇宙，只是在其中加进物质后才发现宇宙将会膨胀，所以他得到一个膨胀宇宙模型是出于偶然。而弗里德曼才是第一个懂得，膨胀是对真实宇宙的相对论描述的必要组成部分，必须从一开始就吸收进宇宙模型之中。他的生命历程虽短，其间却充满变数。他 1888 年生于圣彼得堡，从 1906 年到 1910 年在该城市的大学学数学，后来成了该大学的数学教师，第一次世界大战期间在俄国空军中服役，1917 年革命后成为彼尔姆大学的正教授，然后又于 1920 年返回圣彼得堡到科学院做

研究，1925 年去世时那座城市已改名为列宁格勒。

弗里德曼的研究兴趣原先集中于地球科学，即地磁学、流体力学和气象学。但作为一个有才干的数学家他又喜爱上了爱因斯坦的工作，并于 1922 年发表了他得到的广义相对论宇宙学方程的解。这个解的两个关键特征对现代宇宙学十分重要。第一，弗里德曼从一开始就意识到，方程有一族解，并不是如爱因斯坦所希望的那样只有唯一解，而是有一组不同的解，其中每一个解都描述一个不同类型的宇宙。第二，弗里德曼把膨胀作为一个应有部分包含到了自己的模型中。这在某种程度上是重现了克利福德在 19 世纪 80 年代的思想，即空间可能是均匀地弯曲的，就像肥皂泡的球形表面，但是曲率可以随时间改变，也许是随着“泡”的膨胀而减小。弗里德曼的模型包括好几个变种，[5]有的是泡会永远膨胀下去；另一些是引力会战胜膨胀，泡膨胀到一个有限尺度后又收缩回来；还有的变种带有宇宙常数，其另一种选择，也是今天的优先选择，就是把宇宙常数设为零。[6]但在所有这些模型中至少都有一个时期，其间整个宇宙是以这样一种方式膨胀，即退行速度与距离成正比。

弗里德曼的工作还有一点，虽不能过于强调，但也值得提及。膨胀宇宙（我们这一个或是别的）中的红移不是由于星系穿越空间的相互分离运动，而是由于星系之间的空间本身在像橡皮布一样伸展，空间（或者更准确地说是时空）在膨胀，星系骑在上面被携带着跑开。

弗里德曼的工作发表在一家知名的、被广泛阅读的刊物上，却不知何故没人理睬。爱因斯坦是在弗里德曼的一位同事访问柏林时知道这项工作的，他还在一封给弗里德曼的短笺里肯定了该工作的正确性，但即使是他爱因斯坦也没有意识到该工作与真实宇宙的关联。20 世纪 20 年代时数学家几乎不怎么接触天文学家，天文学家也根本不

注意数学上的新进展，欧洲和美洲的科学界也远比今天分离。所以欧洲的新数学成果就不能很快地与正在美国进行的新天文观测结合起来。是否还有理论家们对这位更以气象学工作知名的俄国数学家抱有偏见的因素呢？不管受冷落的原因究竟是什么，弗里德曼于1925年黯然辞世了，而具有讽刺意味的是，他的死可能正是由他对气象学的兴趣造成的。许多官方传记都说他死于伤寒。但是据宇宙学家伽莫夫(George Gamow)说，弗里德曼的死因是肺炎，是在一次气象气球飞行中受寒而得的。伽莫夫这位下一代宇宙学家的代表人物应该是了解情况的，因为他是弗里德曼的学生，并且在20世纪30年代中期才移居美国。[7]

于是就只得留待下一个人再用与弗里德曼同样的方式解出爱因斯坦方程(是相当独立地，并不知道弗里德曼的工作)以实现认识上的突破，把这些解接受为宇宙学家探索宇宙性质的有效工具。那个人就是勒梅特(Georges Lemaître)，一位比利时宇宙学家。他的有关论文原先是于1927年发表在一家不知名的比利时刊物上，因而没有受到什么注意。但紧随红移—距离关系发布之后，那位无处不在的爱丁顿得知了勒梅特的论文，并安排翻译成英文于1931年刊登在《皇家天文学会月刊》上。[8]如果有任何人当得起“大爆炸之父”的称号，那就是勒梅特。这称号在许多年里成了一个令人敬畏的双关语，因为勒梅特不仅是一个宇宙学家和数学家，还是一名神父。

勒梅特生于1894年，原先学的是土木工程，第一次世界大战时在比利时炮兵部队当军官，战后到卢万大学学习，1920年毕业后进了神学院，1923年被委任为罗马天主教神父。然后他在剑桥同爱丁顿一起工作了一年，又去美国在哈佛和麻省理工学院待了一年，再回到卢万大学，1927年成了天文学教授，一直干到退休，1966年去世。他在漫长的学术生涯里持续地发展着自己的宇宙学思想，并且活着看到自己

的许多思想在宇宙学主流中得到体现。这些思想中最重要的就是大爆炸，尽管他自己没有给出这个名称。勒梅特的解与弗里德曼所得到的基本一样，但他终其一生都宁愿保留那个连爱因斯坦自己都在 20 世纪 30 年代放弃了的宇宙常数。更与弗里德曼或是以前任何人不同的是，勒梅特着手处理了那些方程能揭示出有关宇宙起源的什么这一问题。

勒梅特清楚地知道当时进行的星系红移观测的一些情况，这也与弗里德曼不同。在 1927 年的论文里他认识到星系可以作为量度宇宙膨胀的“检验粒子”，并且在没有什么参考资料的情况下给出了红移—距离关系中比例常数的值(即后来所称的哈勃常数)。他给的值与稍后哈勃发表的值竟如此接近，以致一位当代宇宙学家认为“这两者之间必定有某种联系”。[9]勒梅特把观测和广义相对论两方面贯通了。既然星系现在相隔很远并且在继续远离，那就必定在过去离得较近。如果向过去回顾到足够早的时候，星系之间就会没有空间。在更早的时候就会是恒星之间也没有空间，再早就会是原子之间、甚至原子核之间都没有空间。这就是勒梅特的大胆想象所能溯及的极限了。他设想，现在宇宙中的全部物质都曾集中在一个大约只有太阳 30 倍大的球里，并把这个球称作“原初原子”。这个原子后来向外爆炸，炸出的碎片就成为我们知道的原子、恒星和星系，而星系的相互分离就是由于宇宙的膨胀。这个过程可与作为原子弹能源的核裂变相比拟，即不稳定的放射性原子核自发分裂成碎片各自分散。这个简单的思想自 20 世纪 30 年代以来已被大大发展和修改了。但是现代宇宙学的核心仍然是如勒梅特所首先提出的，我们的宇宙是诞生于一种超密状态，我们在这个膨胀的宇宙里所看到的一切都是从那种状态里产生出来的。随着哈勃和赫马森表明宇宙在膨胀的观测，随着勒梅特关于原初原子的论文于20 世纪 30 年代初被印成英文，现代宇宙学起步，

并向前飞奔了。

当然，在20世纪20年代还有其他先驱者对相对论宇宙学的建立作出了重要贡献。美国数学家罗伯逊(Howard Robertson)扩展了德西特的工作，在宇宙学的数学基础方面很有功绩，的确值得一提。他和英国同行沃克(Arthur Walker)一起于1935年找到了一种对均匀各向同性时空的描述，即所谓罗伯逊—沃克度规。该度规所描述的宇宙有着均匀弯曲的空间和一个对所有随宇宙膨胀一起运动的观测者都相同的宇宙时。这种理想化的宇宙，即罗伯逊—沃克模型，今天正被广泛讲述。但是关键的理论进展是来自4个人，并且正是在斯里弗、哈勃和赫马森积累宇宙膨胀证据的那10年左右取得的。这4位“宇宙学之父”就是爱因斯坦、德西特、弗里德曼和勒梅特。也许还可以加上爱丁顿，他像一位仁慈的教父，当幼儿在走向成熟的路上蹒跚起步时给予了扶助。20世纪30年代初，理论和观测已经以一种非凡的方式结合起来，坚定地朝着大爆炸的方向前进。又过了10年，曾是弗里德曼学生的伽莫夫发展出一套完整的新宇宙学。与此同时，各种不同的替代模型也纷纷涌现。除了一些在不同时候由个别宇宙学家所主张的古怪变种外，方程的解，即弗里德曼和勒梅特在20世纪20年代所找出的解，被归入3个主要类型，而这3个类型都有着有限的年龄。

宇宙的年龄

新宇宙模型在20世纪30年代没有引起科学界震动的一个原因，是从当时对哈勃和赫马森的红移资料所作的最好解释中得出的时间尺度问题。红移给出了对相隔一定距离的星系分离得有多快的量度。[10]如果作最简单的可能假设，即宇宙自大爆炸以来一直以同一速率膨胀，那就很容易计算出自大爆炸至今已经过去的时间，即“宇宙的年

龄”，如果你喜欢这样说的话。20 世纪 30 年代的天文学家这样做时却遇到了意外困难。退行速度等于哈勃常数 H 乘以两星系间的距离。如果膨胀总是以同一速率进行，则自膨胀开始以来的时间，即自两个星系相接触以来的时间，就正是 $1/H$。用哈勃自己得出的哈勃常数值，却发觉宇宙年龄只有约 20 亿年。

这可就使人窘迫了，因为已经有很好的证据表明地球和恒星(包括太阳)都要更古老。在此后 20 来年里事情变得更糟，因为各种不同的技术都清楚地指出，我们在宇宙中所见到的东西绝大多数都比那个简单估计的“宇宙年龄”要更老。地质学的证据，对地球上(还有月球上)和陨石中的物质样本里放射性原子残余的测量，都给出太阳系的年龄至少有 40 亿年，可能更接近于 46 亿年。随着 20 世纪 30 年代核物理的发展，天文学家(爱丁顿是其中突出的一位)开始找出了是什么使太阳和恒星长期维持高温，并推断许多恒星和星系已经存在了不只是 40 亿年而是 100 亿年或更长时间。

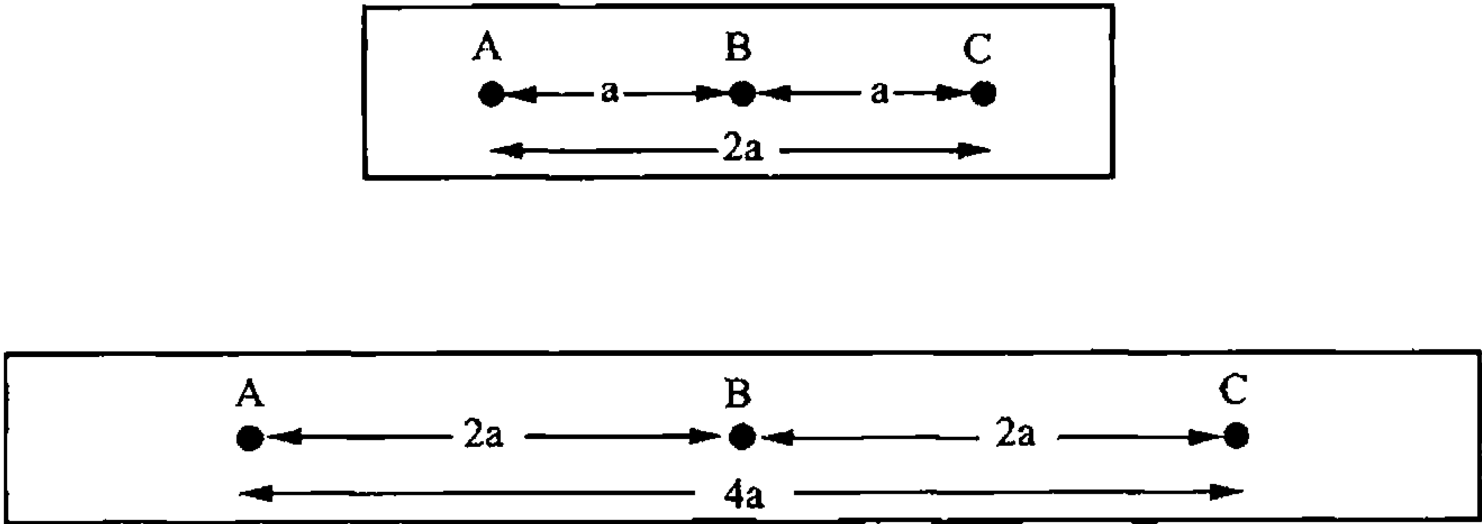


图 4.3　时空膨胀就像拉伸一块橡皮。“星系” A、B 和 C 并不是越过它们之间的空间运动。但当空间膨胀到使 A 和 B 之间的距离加倍时，其他任何一对星系之间的距离也加倍，包括 A 和 C。从这个宇宙中的每一个星系看去，其他每个星系都在以与距离成正比的速率退行。例如，由于 C 与 A 的距离是 B 与 A 的两倍，当所有的距离都加倍(即标度因子加倍)时，就仿佛是 C “离开” A 的速度是 B “离开” A 的速度的两倍。

一种绕过困难的途径是构造更复杂的宇宙模型，留出余地使一切天体得以存在。毕竟方程式中并没有什么东西要求宇宙必定自大爆炸

以来就以不变速率膨胀。勒梅特赞成这样一种修改，即宇宙仍起始于原初原子，但宇宙常数比爱因斯坦静止模型中的稍大一点。这样的宇宙开始时膨胀得相当快，然后就慢下来变成一种几乎静止的状态并停留一段时间，接着又重新膨胀。通过恰当地选择宇宙常数的精确值，勒梅特模型可以做到在准静态需要停多久就停多久，以给出足够的时间让恒星和星系形成并演化。

爱丁顿有另外一个办法。他认为原初原子和大爆炸的思想有着“缺乏美感的突然性”，并提出宇宙可能从开始(不管这“开始”是何意)就处于静止状态，一直到几十亿年前某种扰动使它开始膨胀。这种想法除了提出者外几乎无人拥护。[11]大多数天文学家认为那是人为地生造出一个不自然的模子，就像爱因斯坦当初强迫他的宇宙学方程去符合静止宇宙图像一样。[12]结果再次是，错的是天文学家的宇宙图像，而不是宇宙学方程原来的更简单的形式。

如果宇宙的年龄的确必须是100亿年或更长，而原来对膨胀的简单解释又是对的，那就必定是采用的哈勃常数值太大了。既然宇宙年龄是$1/H$，那么比如说将H减半，得出的宇宙年龄就加倍。20世纪50年代初对宇宙时间尺度就作了这样一种戏剧性的修正。

哈勃常数是通过对遥远星系红移和距离的测量而得出的。对红移从来没有什么疑问。但是宇宙距离尺度的根基就很不可靠，先用三角法直接测出很少几颗造父变星的距离，然后就用作测量宇宙距离的标尺。海尔留下的大望远镜，20世纪40年代一位德国出生的天文学家来到洛杉矶，还有美国参与第二次世界大战，这三者的结合使得银河系外所有天体的距离尺度(从而对宇宙年龄的估计)被修正得猛然大增。

巴德(Walter Baade)1893年出生于德国的希罗廷豪森，是一名中

学教师之子，沿着正规的教育途径直到1919年在格丁根大学获得博士学位。在那个观测和理论双方都开始揭示我们宇宙真实本质的20世纪20年代，巴德在属于汉堡大学的贝格多夫天文台工作。但是德国政治气候的变化使他像许多同时代人一样于1931年来到美国，在威尔逊山和帕洛玛天文台工作了27年，1958年才回到德国，1960年在格丁根去世。在旅居加利福尼亚的中期，巴德经历了一生中又一个戏剧性情节。作为一个德国公民，当美国参战时他被认为不适合直接参加与战争有关的工作，所以在大多数天文同行都被征调去搞军事研究时，他却孤独地也应该说是满意地留在天文台，因为可以无限制地使用当时世界上最大的100英寸(2.54米)望远镜。那时洛杉矶实行战时灯火管制，也没有别人来同他争观测时间，巴德在1943年把望远镜使用到了最大限度，拍出的照片里仙女座星系的内部区域也分解成了单个的恒星光点，而以前哈勃对内区只看到模糊一片。巴德从观测结果中看出，在这个近邻星系里有两类大不相同的恒星。第一类被他称为星族Ⅰ，是年轻恒星，其中许多都是高温和蓝色的，存在于旋臂之中。第二类是在星系中心部分和星系晕中的球状星团里发现的，是年老的恒星，基本上都较冷和发红，他称为星族Ⅱ。我们现在知道这是包括银河系在内的所有旋涡星系的一个典型特征。星族Ⅰ恒星年轻而星族Ⅱ恒星古老，两者之间还有其他重要差别。

当洛杉矶重见光明、同事们也从战时岗位上返回时，巴德仍能继续这项研究，因为200英寸(5.08米)望远镜于1948年启用了。它的大口径超额补偿了城市灯光造成的观测条件恶化。巴德很快就查明，仙女座星系的每个星族都有自己类型的造父变星。两个星族的造父变星都有很确定的周期—光度关系(即周光关系)，但是这两个关系互不相同。哈勃所用的周光关系对星族Ⅱ造父变星是对的，银河系晕里

的造父变星就属此类。但是这个关系被用到了仙女座星系里更高温更亮的蓝色星族Ⅰ造父变星上，而忽视了此类变星有不同的周光关系。星族Ⅰ造父变星比星族Ⅱ中的要明亮得多(这也就是哈勃能分辨出它们的原因)，当巴德用正确的周光关系来重新计算仙女座星系的距离时，他得出了2 000 000光年而不是哈勃所估算的800 000光年。仙女座星系比哈勃认为的更明亮也更遥远。

由于仙女座星系的距离是哈勃估算宇宙尺度的决定性步骤，巴德就一举把所有河外星系的距离翻了一倍还多，把哈勃常数减小到不足以前的一半。既然河外星系比原来所想的还要遥远得多，它们也就必定比原来所算出的要大得多，才能表现得如地球上望远镜里看到的那么大。它们必定与银河系的尺度大致一样，其中有些甚至更大。银河系不再是宇宙中的大个子，而只是很平常的一员，大小与别的星系相近。报纸上把巴德的结果作为头条新闻，大肆宣扬“宇宙的尺度被加倍了”。但对宇宙学家来说更重要的是算出的宇宙年龄也加了一倍还多，由20亿年变成了50亿年。在20世纪50年代初，宇宙看来至少比地球和太阳系要古老。

在后来的30年中，对宇宙距离尺度和年龄的估算几乎是在不断地随着新的观测而改变。巴德的结果被证明显然只是下限。美国出生的天文学家桑德奇(Allen Sandage)于1952年加入了海尔天文台的队伍，他从用200英寸(5.08米)望远镜继续进行的观测中得出，宇宙年龄可能长达200亿年。所有这些结果都仍有一个不确定范围，这是因为那根把我们在一个普通星系边远地区的观测结果连接到宇宙深处的距离估算链条还嫌脆弱。[13]但是，今天即使有的话也只是很少天文学家会反对这样一个估计，宇宙的年龄即自大爆炸以来的时间是在130亿到200亿年之间。这里仍有相当大的余地来容纳不同意见。有的证

据显示，已知最老的恒星有大约200亿年的高龄，但是一些宇宙动力学研究却给出150亿年或更低的年龄。这个争论在试图确定膨胀宇宙的终极命运时是很重要的，但在很大程度上与对大爆炸的探索无关。20世纪30年代得出的宇宙年龄与简单的大爆炸模型明显抵触，而今天的不同意见则是对精细调节的讨论。总的说来，对哈勃常数的观测和计算与用弗里德曼—勒梅特方程构造的最简单类型的模型宇宙符合得很好。

宇宙的选择

我们的宇宙在膨胀，从前比现在要更密集。这个认识的确立不仅是由于把遥远星系的红移解释为膨胀效应，而且还通过对很高红移的射电星系计数并与在低红移处等价空间体积里的射电星系数进行比较。由于光要花费时间跨越(或穿过)空间而到达地球，我们看到的宇宙的遥远部分(即高红移处)其实是很久以前的样子。如果一个星系的红移很大，以至于来自它的光在旅途中用了50亿年，那么我们所看到的这个星系其实是它在50亿年前的样子。所以已经有直接证据表明宇宙不但在膨胀并且来自某种高密态。简单的静态宇宙学模型被排除了。这还意味着能够排除方程所允许的所有简单收缩模型。可能与真实宇宙有关联的就只有那些至少包含有一段时间膨胀的模型。这就仍为一些相当奇怪的模型留有余地，比如说，宇宙起始于某种很低密度的状态，用很长时间收缩并减慢下来，然后再膨胀。但是如在下一章里要讲的，由于现在已经有了关于宇宙是起始于很密很热状态的很好证据，我将不再提那些古怪模型。(当然，如果刚才说的那个模型里的宇宙收缩得很厉害，达到了极高密度和温度的状态，然后再膨胀，那么从人类观测的角度看来这也就成了一个大爆炸模型。)

宇宙从中产生的那个极密集极高温状态通常被称作大爆炸。好像是霍伊尔(Fred Hoyle)在1950年发表的一篇科学论文里把这个词引入了天文学。许多天文学家不喜欢这个词，因为它会使人误解。它给人的印象是在真空中发生了一次爆炸，就像一枚巨大的鞭炮或是一颗放大的核弹。但既然是空间本身也在膨胀，物质是被膨胀的空间携带着，那么炸弹比喻是说不通的。即使在宇宙很密集时，它也很平稳。并没有压强差来使它膨胀，也没有声波来使“爆炸”可以听得见。膨胀是一个平稳事件，一直持续到今天。但是尽管咬文嚼字者正确地表示不满，我们现在可已被大爆炸这个名称钉住了，我也就不想逆潮流而动。

可能的宇宙种类取决于时空的曲率。正曲率可以给出一个闭合面，类似于球面，但是具有正曲率的面也可以不闭合而延伸到无穷远。也许这更容易借助于线即一维情况来理解。一条线可以自己闭合，如一个圆，于是它就有一定的尺度，即圆的周长。但它也可以是开放的，如双曲线，从顶点向两侧延伸到无穷远，并有一个特征形状。我们所生活的宇宙可能是闭合的，即是圆周(或球面)的对应体，其广度是有限的。它也可能是开放的，像一条双曲线或一个双曲面大碗那样，能延伸到无穷远并能容纳无限量的物质。决定弯曲状况并从而决定宇宙究竟是开放的还是闭合的，是宇宙中物质的总量，因为使时空弯曲的引力是由物质产生的。宇宙的开闭对确定其终极命运很重要，但这几乎与对大爆炸的探索无关，因为开放和闭合的宇宙都是以同样的方式起始的。

负曲率则对应于只能开放的曲面，如鞍形面。这种情况是否出现也取决于宇宙中有多少物质(更准确地说，是取决于宇宙演化进程中任一时间即任一“宇宙时代”的物质密度)。爱因斯坦的宇宙常数又

提供了更多的变化余地，但我将只提到这些额外可能性中的一种。

能对所有这一切取得某种贯通认识的最好方式，是看看宇宙学家所称的“标度因子”的性质以及它如何随一个特定模型宇宙的演化而改变。标度因子可被认为是一对选定星系之间的间隔，它随着宇宙的膨胀而增大，用字母 R 表示。当 R 加倍时，任何一对星系之间的距离也加倍，依此类推。将真实宇宙的观测行为即其膨胀反推到所有一切都挤在一个点上，即 R 从零开始，也就是说宇宙从无限密集的状态诞生。如果画一张标度因子 R 随大爆炸以来的时间 t 变化的图，就能看出各种主要类型的模型宇宙如何互不相同。重要的是空间弯曲的程度，以及空间(而非时空)曲率如何随时间变化。在所有情况里(假定宇宙常数为零)宇宙膨胀都随 t 增大而减慢。这很容易由引力对宇宙膨胀的阻碍作用来理解。而减慢的速率则取决于宇宙中有多少物质，物质的量又决定着宇宙是开放还是闭合。

宇宙膨胀随时间流逝而减慢，所以 R 在一段同样长但较晚的时间里所增大的量就较小。这种减慢过程在闭合宇宙里发生得最快。对所有正曲率宇宙，R 线总是确定地向右方弯曲，最后会弯到变成描述某个收缩而不是膨胀的宇宙。这种宇宙中的物质是如此之多，以至于引力最终能使它再收缩。收缩的速率当然会越来越大，这个宇宙就会冲回到无限大密度(正如其由之产生的大爆炸)的状态。这样一个闭合宇宙的空间也像球面一样闭合，曲率随时间的变化相当于球面先长大再缩小。在一个有着负曲率的宇宙里，尽管膨胀总在减慢但将永远进行下去，宇宙就像一个鞍形面那样是无限伸展的。这样一个宇宙的空间被称为双曲型的，这种比拟方式与将闭合宇宙的空间称为球型是一样的。[14]闭合宇宙模型和开放宇宙模型都各有许多亚种，各自组成自己的家族，而爱因斯坦—德西特宇宙是正好处在正曲率模型和负曲率

模型中间的特殊情况。

既然宇宙的诞生已被定名为大爆炸，宇宙学家至少对它的可能结局的命名就能一致了。收缩宇宙的结局是“大坍聚”，而永远膨胀的结局叫什么呢？按艾略特(T. S. Eliot)很有逻辑性的说法应是爆炸的反面——“这就是世界结束的方式，不是以一声爆炸而是以一声呜咽。”所以就有两种基本选择，爆炸—坍聚宇宙和爆炸—呜咽宇宙。其余的一切就都只是微调。[15]

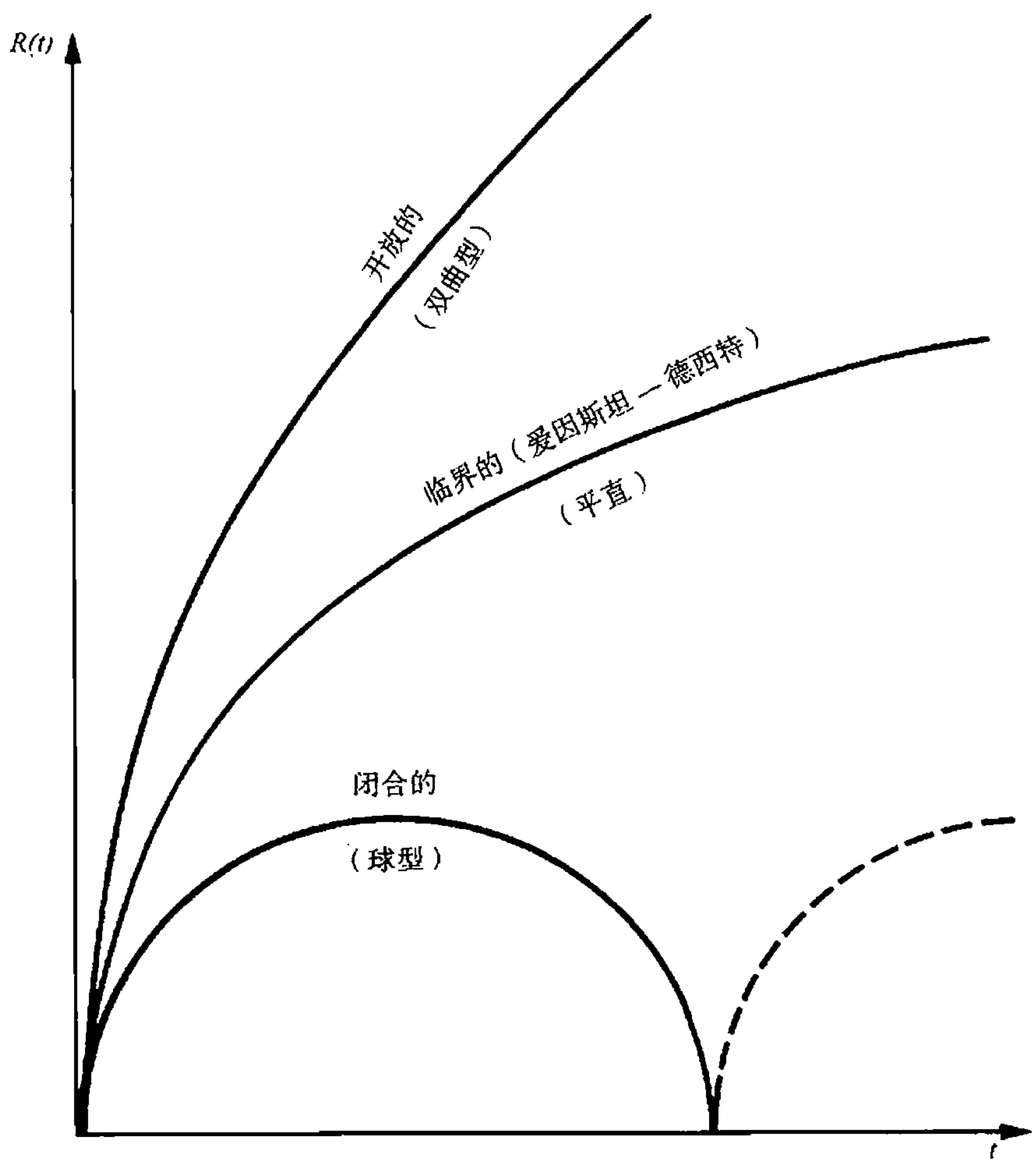


图 4.4　图 4.2 中描述的三种空间几何对应着三种类型的宇宙，全都从致密状态开始并至少在初期是随时间流逝而膨胀。R 是宇宙的标度因子，t 表示时间。开放宇宙和平直宇宙都永远膨胀；闭合宇宙最后会收缩，并可能随后又进入膨胀和收缩的新循环。

图 4.4 所示为这些不同类型的宇宙如何演化。该图还表明为什么

用哈勃的参数 H 对真实宇宙年龄的估计必定只是一个近似值。把宇宙年龄定为 $1/H$，就意味着假定膨胀总是以同一速率进行，实际上就是画出 $R(t)$ 曲线的一条切线并反向延长到与 t 轴相交，即找出 R 为零的时刻。如果正曲率空间的膨胀是减慢的，那么 $1/H$ 这个估计就总是会比宇宙的真实年龄更长。这就使哈勃宇宙年龄与对地球和恒星估测的年龄之间的冲突在 20 世纪 40 年代变得更使人为难，因为在哈勃年龄中总需要有合理的余量来保证宇宙已足够老，使恒星和行星得以形成。当伽莫夫于 20 世纪 40 年代中期给出第一个详细的大爆炸模型时，它没有马上被看作宇宙学家所祈求的答案并不足怪。然而，这个模型经受住了时间的考验。

大与小

勒梅特的宇宙模型首次把关于小世界和大世界的已有的最好思想合并到对现实的同一个描述之中。这在当时是很勇敢的一步，但在 20 世纪 80 年代看来，粒子物理学和宇宙学的结合是认识我们生活于其中的宇宙的唯一途径。20 世纪 20 年代后期，当天文学家刚刚开始认识可见宇宙中的物质分布时，物理学家也刚刚开始认识原子之中物质的分布。19 世纪后期所发现的放射性给了物理学家一个探测原子内部的工具。他们用天然放射性原子中产生的所谓 α 粒子作为小子弹去射击晶体或金属箔中的原子。以新西兰出生的物理学家卢瑟福(Ernest Rutherford)为首的英国曼彻斯特大学的研究者运用这种技术发现，绝大多数 α 粒子都直接穿透薄金属箔靶子，但偶然有个别粒子会几乎沿原路反弹回来。卢瑟福于 1911 年对这个现象作了解释，给出了今天在中学里讲授的原子基本模型。[16]

卢瑟福意识到，原子中的绝大部分物质必定集中在其中央一个很

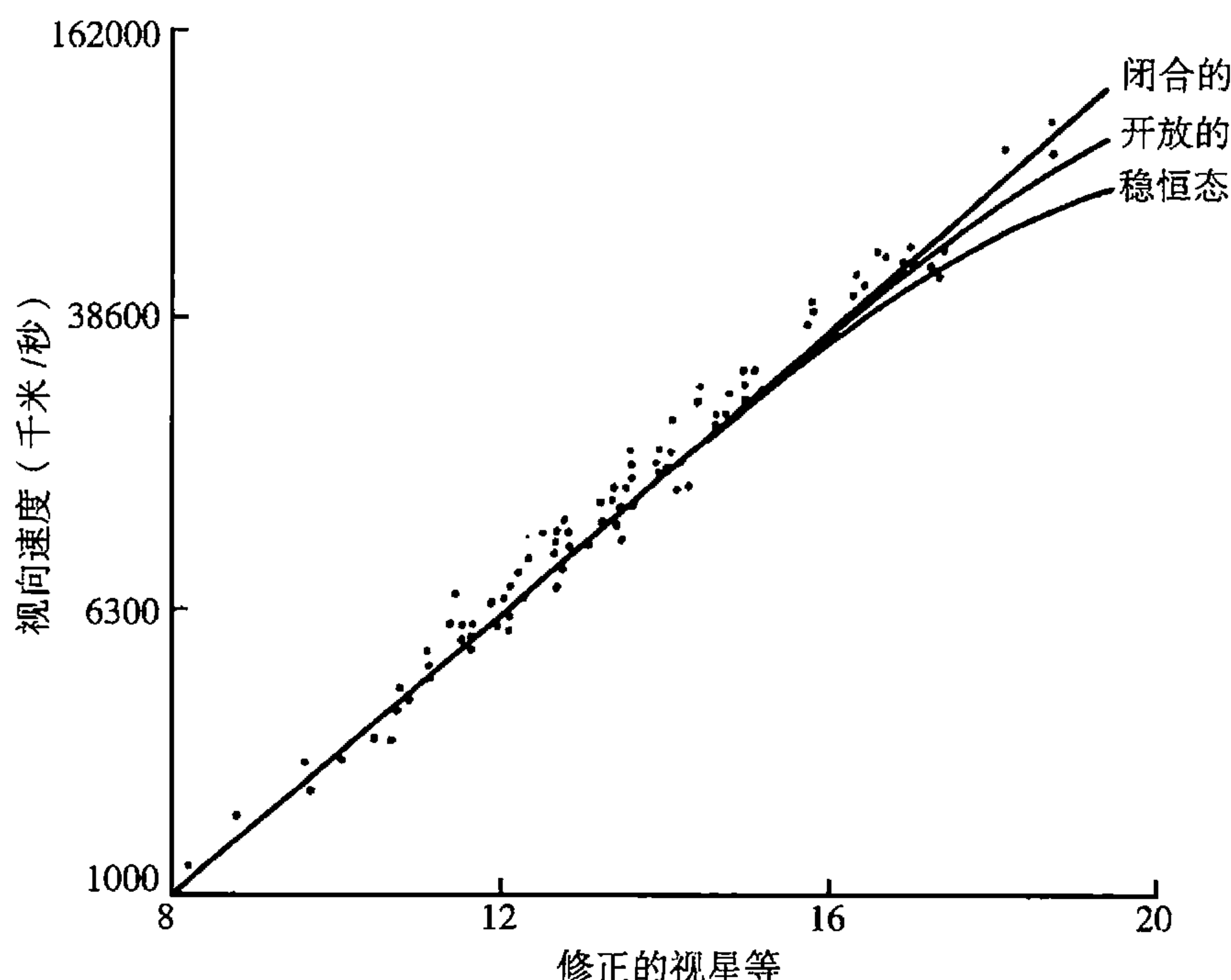


图 4.5 对遥远星系的观测原则上可以揭示出我们的宇宙是何种类型。我们所看到的遥远星系是它们在宇宙更年轻也膨胀得更快时的样子，那时的“哈勃常数”值是不同的。所以哈勃图上的直线应该稍稍弯曲，弯曲量则指示出宇宙膨胀减慢的速率。可惜现有望远镜的观测范围还不能提供足够的证据来决定图 4.4 中的哪一个模型最值得下赌注。虽然稳恒态宇宙看来可以排除，但凭（由桑德奇用多年时间收集的）现有观测仍无法区分我们的宇宙是开放还是闭合。这样看来，我们的宇宙可能很接近于那个特殊的中间情况，即平直时空。此图中视星等实际上对应着距离，而视向速度对应着红移，所以这是图 3.1 哈勃图的现代的、大大扩展了的版本。

小区域里，他称之为核，核的周围是电子的云。来自放射性原子的 α 粒子实际上是原子核的碎片。当这种粒子打到原子的电子云时，它一掠而过，几乎不受影响。电子带负电荷，而原子整体是电中性的。所以原子中的正电荷必定像其质量一样集中在核里。α 粒子也带正电荷。当一个 α 粒子迎头撞上一个原子核时，同种电荷间的斥力会使该粒子停下并将它推回去。后来的实验证实卢瑟福的原子图像基本上是正确的。绝大部分质量和所有正电荷都集中在核里，核的大小仅约为原子的十万分之一。占据其余空间的是很轻并带负电的稀薄电子云。核的直径是 10^{-13} 厘米量级，而原子直径是在 10^{-8} 厘米量级。粗

略地讲，这个比例就像一粒沙放在卡内基大厅中央，空荡荡的大厅是“原子”，那粒沙就是“核”。

核里带正电荷的粒子叫做质子。它所带电荷的量与电子电荷精确相等，但符号相反。它的质量大约是电子的2000倍。在卢瑟福最简单形式的模型里，原子中除电子和质子外别无它物，二者数量相等，质子集中在核内，纵然它们带有同种电荷理应相互排斥(就此而言同种电荷的行为与同种磁极的行为相同)亦然。必定有另外一种力，在很小的范围起作用，能克服电力而把核粘牢，本书第三篇将详述此事。但在卢瑟福原子模型提出后的20年里，物理学家中有一种猜疑在增长，即应该有另外一种粒子，一种质子的补足物，质量与质子一样，但不带电荷。核里此种粒子存在的可能作用之一，就是使带正电的质子能聚集在一起而不被电力推开。存在这种很快就被称为中子的粒子，还能解释为什么有些原子化学性质一样而质量却稍有不同。化学性质取决于原子中的电子云，那是给别的原子看的“面孔”。化学性质一样的原子必定有着相同数量的电子，从而也有着相同数量的质子。但它们却可以有不同数量的中子，从而有不同的质量。这样的原子亲属现在叫做同位素。[17]

世界上多种多样的元素全都统一于这个简单方案之中。氢原子最简单，核里只有一个质子，核外一个电子；最常见的一种形式的碳原子是包括我们人体在内的生命物质的基础，其核内有6个质子和6个中子，6个电子的云围绕着核。有的原子核含有比这多得多的粒子(统称为核子)。铁的核里有26个质子，它最常见的一种同位素有30个中子，共56个核子。铀是天然存在的最重的元素之一，在铀235的核里有92个质子和不少于143个中子，这是一种用作核能源的放射性同位素。从很重原子核的裂变可以获得能量，因为原子核可能具

有的最稳定亦即能量最低的状态是铁56。从能量上讲，铁56是在谷底，比它轻的核，包括氧、碳、氦、氢在内，在一侧向上排列；而比它重的核，包括钴、镍、铀、钚等，在另一侧向上排列。正像一只放在谷坡上的球很容易一踢就滚到谷底，而要踢得往上滚就难得多一样，重原子核在适当条件下也会分裂，“滚下坡”而变成更稳定的核，同时释放能量；同样，轻原子核也能聚合在一起而成为更稳定的核，并释放能量。[18]裂变，即勒梅特试图推广到原初原子上的过程，是原子弹的能源机制；而聚变，即伽莫夫运用到他的大爆炸模型中的过程，则是氢弹(或称聚变弹)能量之所由来，氢弹里进行的是氢核聚变成氦核的反应。但在20世纪20年代，这一切还是将来的事。尽管当时就有不少关于中子存在的证据，直到1932年才由查德威克(James Chadwick)做出了证明中子的确存在的实验。查德威克原是卢瑟福的学生，当时在由卢瑟福当主任的剑桥卡文迪什实验室工作。

所以，当勒梅特提出他关于宇宙起源的“原初原子”模型时，没有人真正知道真实的原子是什么样子。“原初原子”这个名称本身就不恰当，如果用“原初核”就好得多。当我们谈论原子“分裂”或放射性衰变时，真正的意思应是原子核破裂成两个或更多个部分，或是核喷射出一个粒子，比如α粒子，并变成一种较轻元素的核。勒梅特所设想的过程是，一个“核”反复裂变而产生出宇宙中的所有物质。但是，这不能解释(且不说此处不拟涉及的更专业性的问题)为什么恒星和星系中物质的一大半是氢这种最轻最简单的元素，而剩余的一小半里绝大部分又是第二轻的元素氦。对恒星和星系的光谱学研究确凿地证明宇宙是由两种最简单的元素主宰，即核内只有一个质子、核外一个电子伴随的氢，和核内有两个质子和两个中子、核外有两个电子的氦。把原初核全都分裂成这么简单的成分是一件过于困难的

事；伽莫夫想出了一个新主意，即大爆炸可以是从最简单的粒子开始，宇宙中的重元素是后来把更多的质子和中子加到最简单的核里而形成的。毕竟，假如从氢开始，那么一下子就能解释宇宙中一大半原子核的存在。

伽莫夫的宇宙

伽莫夫是一位超常人物，有着无限的想象力，因而能从核物理转到宇宙学然后又进入分子生物学世界。他对这三个科学领域，也是20世纪的三个关键科学领域，都作出了重要贡献，却还能找出时间来给外行人写书，在书里给同行们精心搞点恶作剧，又全面地展现了20世纪中叶的科学世界。可他对拼写或日期之类的日常小事却能力甚差，甚至连简单的算术都做得不好。他1904年生于乌克兰的敖德萨，20世纪30年代中期永久移居美国之后，给朋友写信时总把自己的名字签成"Geo"，他坚定不移的这个缩写读音是"Joe"，所以对许多朋友来说他是"Joe"，直到1968年去世为止。

经历了俄国革命和内战的动荡之后，伽莫夫于1922年进了新罗西斯基大学，但很快又转到列宁格勒大学，在那里待到1928年，得了博士学位，也从弗里德曼本人那里学到了弗里德曼宇宙模型。一有了资格，他立即去了格丁根大学，然后是哥本哈根的理论物理研究所，接着是剑桥的卡文迪什实验室，又返回哥本哈根。他在1928—1931年间访问的这三个科学机构正是当时发生的物理学革命的中心，标志这场革命发生的是量子物理学的出现并开始应用于对原子的认识。伽莫夫的量子物理是从该学科的首创者们那里学来的，正如他的宇宙学是从其首创者之一那里学来的一样。他在访问格丁根期间作出了自己的第一个重大的科学贡献，即用量子理论解释 α 粒子是怎

样从原子核中逃逸出来的。

现在知道，每个 α 粒子由两个质子和两个中子组成，很强的核力克服了质子之间的电斥力而使这 4 个核子组合在一起。它们其实就是氦核，即氦原子的全部两个电子被移去后的剩余物。当一个 α 粒子是在一个很重的原子核里时，它被强核力束缚着。但是如果它是刚刚在核外面，电斥力就会起主宰作用，因为核力的作用范围很小，于是它就被射出。把稳定态是在能量谷底的概念加以扩展，对 α 粒子来说核就像一个死火山口的内部。当 α 粒子处在火山口内深处时，它是在稳定能态；但如果它是在火山口外一点点，那就是在山峰的陡坡上并很快地滚下来。伽莫夫说明了 α 粒子如何能越过那火山口的小峰而从恰在核内移到恰在核外。他对 α 衰变的解释是量子理论对原子核的第一次成功应用。

1931 年伽莫夫被召回苏联，并被委任为科学院的研究员和列宁格勒大学的物理学教授。但他热情奔放的本性和独立的思想难以适应 20 世纪 30 年代斯大林时代的生活，当他于 1933 年获准参加在布鲁塞尔举行的一次学术会议时，就留在国外，去了美国首都华盛顿的乔治 · 华盛顿大学。1934—1956 年他在该大学当物理学教授，然后去了博尔德的科罗拉多大学，直至去世。

伽莫夫对粒子如何从原子核里出来的兴趣又引导他想知道粒子进到核里面去的可能性，即粒子从外面越过山峰而进入强核力控制的区域。他作了这方面的开拓性计算，证明了如果把能量为几十万伏的质子射进原子中，就会引发核反应导致核裂变和 α 衰变。卡文迪什实验室的科克罗夫特(John Cockroft)和瓦尔顿(Ernest Walton)在 1932 年正是这样做的，他们创造出了世界上第一个粒子加速器，用高压电场使质子加速并撞击原子，从而引发了恰如伽莫夫预言的反应。这是

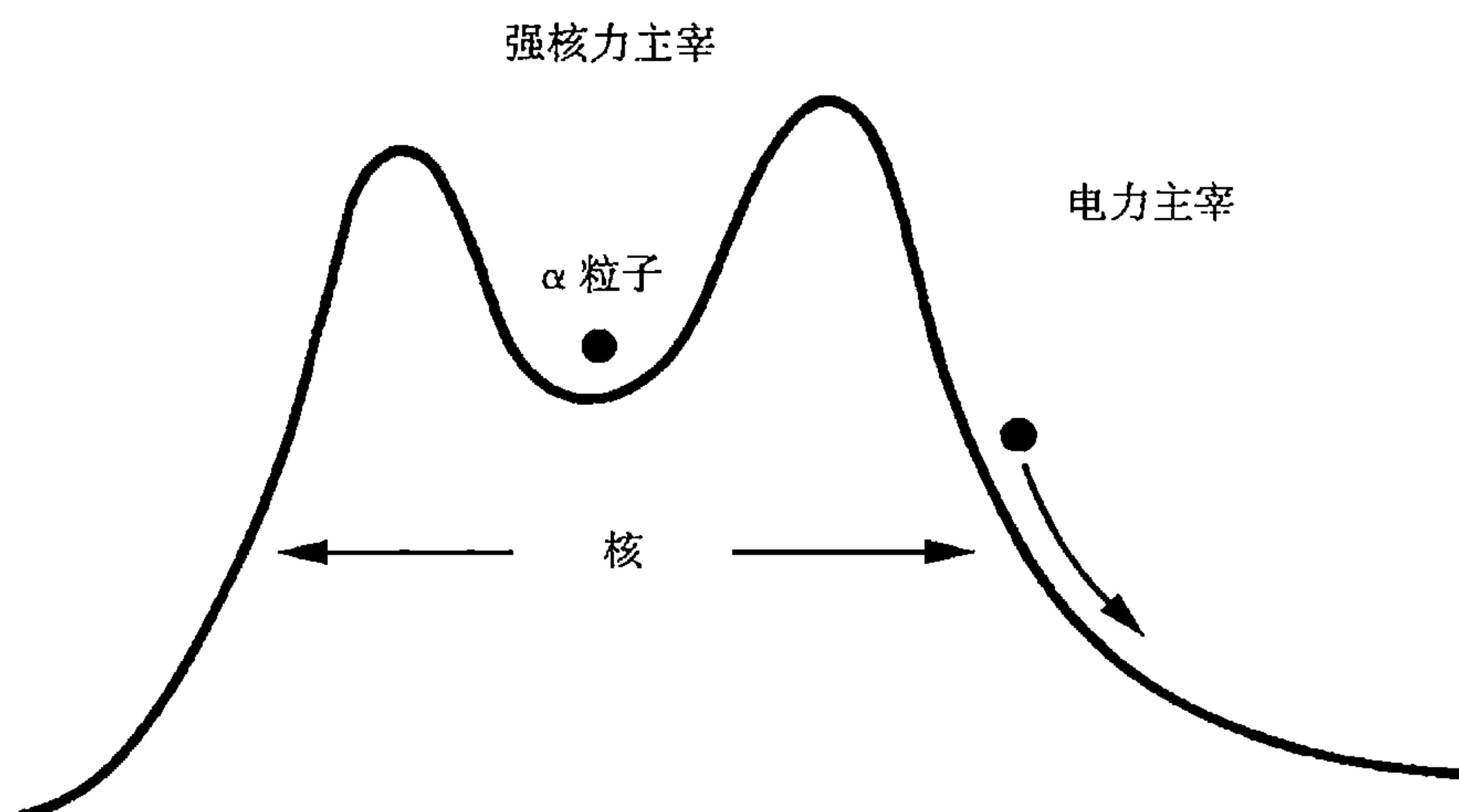


图 4.6　α 粒子在核内被强核力控制。但它若是恰在核外一点点，就会被电力推斥而“滚开”。但是 α 粒子如何能越过核内与核外之间的能量峰呢？伽莫夫对这种 α 衰变过程的解释是量子物理学对原子核的首次成功应用。

第一步，继续走下去将最终导致原子弹和核电站，即从裂变过程获得能量。但这个思想，即克服长程电斥力把质子推进到短程核力控制之下从而粘结到原有的原子核上，也把伽莫夫引向了大爆炸。

中子是伽莫夫宇宙的关键成分。一个中子只要是在稳定的原子核内，它就保持为中子。但若随其自便，则单独的中子会自己衰变，每个中子分裂成一个质子和一个电子。这种衰变发生得相当快，半衰期大约是 13 分钟。[19]所以如果宇宙是从满含中子的高密集态也就是一种中子气开始，那么它就很快能提供出质子和电子，这两种粒子的数量总是相等的，它们又进而组成稳定的原子，于是就不会有一种电荷过剩。

伽莫夫的主意立即给宇宙提供了氢。每个氢原子仅由一个质子和一个电子组成，由异种电荷之间的吸引力维持在一起。允许一个中子衰变，就有了一个裸露的氢核和一个近便的电子，也就为构造出一个原子作好了准备。但是其他种类的原子，即氦和更重元素的原子，又是从何而来的呢？

20世纪40年代，乔治·华盛顿大学的一位名叫阿尔弗(Ralph Alpher)的研究生参加进来了，他被伽莫夫指派去详细研究在大爆炸模型中更复杂的核如何能从氢产生出来(这个过程现在称作核合成)。他们两人认为，这是依靠宇宙最初几分钟内那稠厚的物质“汤”里粒子之间的碰撞。[20]他们的计算表明，相对容易的是一个质子(氢核)和一个中子碰撞并粘在一起形成氘(也叫重氢)核。氘核再次与一个中子碰撞就产生氚核，其中有一个质子和两个中子。但氚核是不稳定的，其中的一个中子会很快分解出一个电子而变成质子。那个核就变成了氦的一种同位素的核，即含有两个质子和一个中子，显然应被叫做氦3。现在需要的是再把一个中子加到在长大的核里，以得到α粒子即氦4原子的核。到目前为止一切顺利。不必担心电子，因为一旦核给造出来了，就很容易从原初浓“汤”的密集粒子里抓到所需要的电子。但是走到这里模型遇上了困难。

氦4核即α粒子特别稳定。它既不愿意碎裂成小部分，也不肯接受新增成分而长成更复杂的核。更糟的是，核里含5个粒子的元素并不天然存在，当这种核在实验室里通过由中子轰击氦4而人工合成时，它立即又分裂成为氦4。伽莫夫和阿尔弗只好绕过这个困难，设想一个氦4可能偶然地同时碰上两个粒子，并将它们都捕获过来而形成一个有6个粒子的核。即使这真能发生，同样的问题又会对含8个粒子的核出现，因为那种核会很快分裂成两个α粒子。[21]随着宇宙从大爆炸的超密态向外膨胀，它很快地变稀疏，到氦形成时那种双重碰撞的机会就很小，而且迅速地变得更小了。在20世纪40年代，尽管由一次捕获两个粒子来越过鸿沟的指望看来未必靠得住，也应看到正是由于对早期宇宙条件和那些核反应发生速率的无知，才使得伽莫夫和阿尔弗把这个猜想用作一个起作用的假设。伽莫夫毕竟告诉了所有

感兴趣的人，他的理论解释了宇宙中所有的氢和所有的氦是从哪里来的，而这两样已经占了恒星和星系中可见物质的99%以上。尽管他的理论不能恰当地解释重元素（天文学对除氢和氦以外的任何元素都这样称呼）的合成，那也只是不到1%的问题。

关于核捕获中子或质子能力的详细计算（计算出来的量叫做捕获截面）就成为阿尔弗于1948年提交的博士论文的基础。这结果显然应该让更多的人知道，于是阿尔弗和伽莫夫写了一篇文章准备提交《物理学评论》发表。这时候伽莫夫那喜欢嬉闹的本性又占了上风，于是就有了他那个最著名的科学玩笑。他后来在《宇宙的创生》一书中写道[22]："按希腊字母顺序，一篇文章只由阿尔弗和伽莫夫署名是不公正的，所以贝特（Hans A. Bethe）博士的名字在准备付印的文稿里加到了中间。贝特博士也收到了一份文稿，没有表示反对。"于是这篇标志着现代大爆炸模型诞生的经典性论文就以阿尔弗、贝特和伽莫夫三人的名义于1948年4月1日发表了，这个巧合更使伽莫夫很高兴。*直到今天，它都以"αβγ"论文而知名**，这倒是恰当地反映了它论述的是事情的开端这一事实，也反映了粒子物理学对宇宙学的重要性[α粒子已经讲过了，β粒子是电子的另一个名称，γ射线是一种强电磁辐射脉冲（一种高能光子）的名称]。

就在同一年，即1948年，霍伊尔、戈尔德（Tommy Gold）和邦迪也提出了他们的稳恒态膨胀宇宙模型。这两个对立模型在专业人士中激起了辩论，辩论贯穿了整个20世纪50年代并延续到60年代，霍伊尔和伽莫夫分别为两派之首，展开友好竞争。有趣的是，将要由霍伊尔来指出如何解决伽莫夫模型的最大困难，因为他是恒星合成重元

---

* 4月1日是西方风俗中的愚人节。——译者

** αβγ分别是三位作者姓氏的谐音。——译者

素过程的发现者之一，而大爆炸只要能完成制造氦的最初工作就够了。但是在整个故事里还有一个更大的嘲弄，它是科学史上误失最重要机会的事例之一，并且着重表明了连宇宙学家当时都没能认真看待他们自己的方程。

两个问题

在那些日子里，宇宙学在很大程度上是一种游戏。竞争的模型被提出来并互相检验，几乎是一种抽象的数学斗智，没人想到这些模型可能有一个会是我们宇宙的正确数学描述。即使是伽莫夫，他爱他的宇宙理论就像是自己的儿子一样，也落入了这个陷阱。

在大爆炸中仅造出氦(现在且别操心重元素)所需要的条件就包括极高密度和极高温度这二者。虽然也可以设想一种冷中子汤从极高密度态向外膨胀，相当简单的计算却表明这种冷中子汤会很快地几乎全部转变成氦。只有在热大爆炸中大部分物质才能保持为氢，而且(这很奇怪但仅是此类奇事之一)在创世瞬间之后几秒钟模型宇宙的密度究竟是多少并不会造成太大差别。只要宇宙是热的，结果总是大约 1/3 的物质变成氦，其余的保持为氢，一直到随着宇宙的演化而在恒星里有新过程启动。

大部分氢被阻止变成氦，而宇宙中则密集着大量的高能辐射。这种电磁辐射可以看作是一种叫做光子的粒子。阿尔弗和另一个年轻的研究者赫尔曼(Robert Herman)按照宇宙物质大约 1/3 是氦而其余是氢这一事实来计算宇宙中应该有多少光子，结果是令人惊愕的，每个核子(即每个质子或中子)都对应着 10 亿个光子。辐射即光子是一种能量形式，辐射的密度(即单位体积空间中的辐射能量)可以用温度来表示。阿尔弗和赫尔曼把弗里德曼的解用到宇宙的最初几秒钟，证明

必定有过这样一段时间，其辐射能量密度大于由爱因斯坦的著名公式 $E = mc^2$ 所给出的物质能量密度。伽莫夫宇宙是诞生于一个辐射火球，随着膨胀而迅速冷却，只是在膨胀并冷却到一定程度以后才变成由物质主宰。但是辐射仍在那里，充满着整个宇宙，只不过随着时间进程而变得更稀、更冷、更弱。1948 年，阿尔弗和赫尔曼发表了一篇文章，其中给出现在这种残余辐射的温度必定大约为绝对零度以上 5 度，即 5 开。[23]

伽莫夫在他于 1952 年出版的科普书《宇宙的创生》里对今日宇宙的温度作了一个稍有不同的估计[也许还因为不满于赫尔曼固执地拒绝把名字改成德尔特(Delter)*]。他推出了一个公式，即温度等于 1.5×10^{10} 除以以秒为单位的宇宙年龄的平方根，这就给出约为 50 开的结果。在 20 世纪 50 年代初，他和同事们还得出今日宇宙温度是在 5 开到 50 开的范围内，这要看对宇宙的早期状态和年龄作什么样的假定。今天的粒子物理学家作了更精确的计算，表明伽莫夫公式中的 1.5×10^{10} 应修正成简单的 10^{10}，而且估计的宇宙年龄也增大了，这两方面都使对当今宇宙温度估计值的上限降低。那个公式只是近似的，还有更好的方法来计算宇宙在任何时期的温度。但它仍不失为一个很有用的约略估计量，比如说，它告诉我们宇宙在创生瞬间后 1 秒时的温度是 100 亿度，100 秒时已降到 10 亿度，1 小时时是 1.7 亿度。作为对照，据计算太阳核心的温度约是 1500 万度。

热大爆炸理论就作出了一个清楚的预言，即整个宇宙应该是辐射的大海，其能量相当于几开的温度。这种辐射应该可以在射电波段探测到，而 20 世纪 50 年代初射电天文学正在起步。但是没有一个射电

---

* 希腊字母表上第四个字母 δ 的谐音。——译者

天文学家注意了这个预言并去检验它，而伽莫夫和他的小组又转到了别的研究领域(他自己对揭开DNA遗传密码的问题着了迷)，再也没有复发原来的兴趣，也没有去鼓励或是迫使射电天文学家采取适当行动。这失误是怎么造成的呢？最好的解释是由物理学家温伯格(Steven Weinberg)在他的《最初三分钟》一书中作出的。他写道，在那时候“对物理学家来说认真地看待任何关于早期宇宙的理论都是极其困难的”，“我们的错误不是把我们的理论看得太认真了，而是没有足够认真地看待它们。我们总是难以意识到，我们在桌面上玩的这些数字和方程式竟然当真会与真实世界有关联。”[24]

1956年，当伽莫夫去了科罗拉多而他的小组也就此解散时，热大爆炸模型的早期版本留下了两个问题，对它们的回答将成为继续发展的基础。第一个问题是普遍认识到的，对它作出解答的努力在20世纪50年代后期和60年代初期取得了巨大进展。那就是重元素若不是在大爆炸中产生又是从何而来呢？第二个问题却一直埋藏在文献里而无人注意，到1964年才纯属偶然地做出了回答。那就是关于今日宇宙的背景温度问题。这两个答案都导致了诺贝尔奖的授予，它们与伽莫夫宇宙的结合开创了宇宙学的现代纪元。

注　释：

1. 引自诺思书中第73页。他在死后才发表的作品《精确科学的常识》(*Common Sense of the Exact Sciences*)中更进了一步，猜想空间可以在所有方向上同等地弯曲，但“弯曲程度可以整体地随时间变化”，这是相对论宇宙学的膨胀宇宙模型的非凡先声。见哈里森《宇宙学》(*Cosmology*)155页。

2. *The Measure of the Universe*(London: Oxford University Press, 1965).

3. 这两句话已被广泛引用并且一般都认为是爱丁顿说的。尽管也有可能这话是他从别人那里借用的，但颇像他的口气，也的确出现在他1932年的一篇文章里。此文刊于*Proceedings of the Physical Society*(Vol.44,p.6)。

4. *Eddington*,p.38.

5. 以两篇论文分别于 1922 和 1924 年发表在德国的《物理学期刊》(*Zeitschrift für Physik*)上。

6. 爱因斯坦—德西特模型是弗里德曼方程的一个特殊的、非常简单的解。

7. 见伽莫夫的自传《我的世界线》(*My World Line*, New York: Viking Press, 1970)。这个故事也由哈里森在其《宇宙学》中重述。伽莫夫还将爱因斯坦对弗里德曼工作的最初反应作了一番精细剖析。他说，弗里德曼在自己的工作发表之前先写信给爱因斯坦，但一直没有收到回复，直到一位同事从俄国去柏林访问时见了爱因斯坦，才得到一封“粗暴”的信说那工作是正确的。正是在那位大人物勉强点头之后弗里德曼才决定发表他 1922 年的那篇文章。爱因斯坦的粗暴也许与弗里德曼指出了他 1917 年论文中的一个简单错误有关；爱因斯坦在方程两边都除以宇宙常数，忘记了该常数可以被设为零，而用零作除数是不允许的。众所周知，伽莫夫自己的算术也是靠不住的，所以对这个故事津津乐道。

8. 勒梅特的原始论文刊登于布鲁塞尔科学学会的《年鉴》第 47A 卷第 49 页。爱丁顿对其热情支持，自己先对该论文作了一个评介，写入一篇文章于 1930 年在《月刊》第 90 卷第 668 页发表，然后又将其译成英文发表于 1931 年(第 91 卷第 483 页)。

9. P.J.E. Peebles, *Modern Cosmology*(Princeton, N.J.: Princeton University Press, 1971), p.8。

10. 最好是从任何两个遥远星系的分离，而不是从星系“离开我们”的退行来考虑问题。我们的红移测量只能在地球上进行，而且确实显示出星系在离开我们退行，但这是一种误解。我们不是在什么特殊的位置上，不是在宇宙的什么推斥中心。只是由于整个时空网在膨胀，宇宙中的任何观测者都看到同样的情形。想想一个完全光滑并画有一些小点子的泡，当泡膨胀时，每一个点都离其他任何一个点越来越远，从任何一个点看去都显得是所有其他点子正在离开它。爱因斯坦在他 1917 年的宇宙学论文里提出了一个基本假定，就是宇宙没有一种平均性质来定义其中的一个特殊位置或是特殊方向。这被称为宇宙学原理，它实际上是说我们生活在宇宙的一个普通区域，我们所看到的与在宇宙中任何其他地方看到的，平均说来都完全一样。当然，爱因斯坦从这个基本假定出发，也许只是因为能得出他认为是最简单的模型，而他也想要自己的方程给出简单的模型，至少开始时是这样。随着逐渐地认识到宇宙确乎符合这种最简单的可能描述，宇宙学原理也被提到了今天这种高贵的地位。

时空的均匀膨胀(或均匀收缩，这在数学上是前者的镜像对称)是宇宙唯一的能符合宇宙学原理的动力学形变，这一点惊人地预示着爱因斯坦的假定(或猜想)是说中了宇宙的某种基本特征，特别要记住他提出这个原理时根本不知道什么宇宙膨胀。假如他的思路转到这个方向上，他仅凭宇宙学原理就能预言，如果天文学家有朝一日发现了宇宙中的大尺度运动，他们所发现的就要么是均匀膨胀，要么是均匀收缩，而没有其他可能性。他的方程式也的确作出了这种预言，但他却用宇宙常数来阻止了。

宇宙学原理与马赫原理有密切的哲学关联。后来对宇宙中物质分布和大尺度运动的观测提供了支持这一点的有力而详细的证据，但这属于那种不可能被确切证明为正确的事情。然而，如果宇宙是构造得在不同观测者看来大不相同，那么宇宙学研究也就根本没有多大意义了，因为那就不可能从我们在自己这个小角落里作的观测推导出关于宇宙整体的任何东西。实实在在地说，没有宇宙学原理，就没有宇宙学。

11. 对时间尺度困难的最好、也是哲学上和美学上最合意的解决方案是 20 世纪 40 年

代末由邦迪、戈尔德和霍伊尔三人提议的。他们想出了一个膨胀着的稳恒态宇宙，其中新物质被不断地创造出来，以填补随着时间流逝和宇宙增大而在原有星系之间留下的空隙。这个模型惊人的简单性对许多数学宇宙学家有吸引力，而许多观测天文学家则觉得至少是给他们提供了一个射击的靶子，无论他们是否喜欢模型背后的哲学，也无论他们是否在意它“对”或“错”。他们所能进行的唯一确定的宇宙学检验是看看是否有任何观测会使该模型无效。（要使大爆炸的思想无效当然就难得多，因为几乎任何东西，包括那种长时间的静态行为，都能与某种大爆炸模型相容！）于是稳恒态模型激发了 20 年的研究，它给了观测家可检验的预言，也建议了可以在稳恒态和大爆炸二模型间作判决的观测。通过激发观测家加倍努力，稳恒态模型无疑加快了宇宙学研究的进度。对那些和我一样仍然觉得该模型背后的哲学有吸引力的人来说，不幸的是它被弃诸路旁，因为有不断增加的证据表明确实有过大爆炸。但是，尽管它可能是一个不正确的设想，它却肯定是一个好的设想，因为它能被检验并且促进了科学认识的发展。

12. 有点嘲讽意味的是，爱因斯坦放弃他的宇宙常数主要是由于勒梅特工作的结果，而勒梅特却又保留宇宙常数以得到他认为满意的模型！

13. 20 世纪 50 年代以来，对宇宙的探测不仅是用光学望远镜，还有射电望远镜，以及较晚的送到地球大气以外的 X 射线探测器和其他装置。射电观测对确定宇宙是在演化起了重要作用，但是射电天文学的详细历史已超出本书的范围。

14. 这些关于球型和双曲型空间的讲述可能会使你想知道是否还有别的可能性，比如说椭球型。不能这样类推，现在最好还是集中于双曲型(开放)和球型(闭合)这两类模型的区分。但是宇宙学家的确有时也尝试一下拓扑结构稍有不同的空间，他们称为椭球型空间。这种空间有一些有趣的性质。它的体积只是有相同“半径”的球型空间的一半。这是因为，在球型空间里每一点都有对应的对蹠点(就像地球上的南极和北极)，而在椭球型空间里相对于一个选定点的最远“点”却可以是在空间某一区域的任何地方，这就像赤道上所有的点与北极都有相同距离一样。但即使我们生活于其中的时空是这种椭球型的(并没有这样的证据)，也并不影响为什么它在膨胀和它怎样诞生于大爆炸这个难题。

15. 我们宇宙的结局取何种选择并有什么微调，这是一个重要的话题，但超出了本书范围。也许值得在这里提一下的是，由引力闭合的宇宙就等价于一个黑洞。实际上，它就是一个黑洞，或者换句话说，在 20 世纪 70 年代和 80 年代强烈地吸引着公众的想象力的黑洞，那些完全由爱因斯坦方程描述的被引力闭合的时空区域，的确就是袖珍宇宙。

16. 我的书《薛定谔之猫探秘》对粒子物理所有这些先驱性工作作了完整叙述。这里讲的只是一个梗概，但我希望能说明物理学和宇宙学是怎样并肩发展的。

17. 关于原子层次化学的更完整叙述可以在我的书《双螺旋探秘》中找到。

18. 正如踢球时给球足够多的能量就能够使它滚到坡的高处一样，在聚变过程中加进能量也有可能造出很重的元素。同样，正如球在没人碰它时能停在半山坡上一样，许多原子核即使严格说来是处在比铁 56 更高的能量状态，也相当稳定。这一切对认识元素原先是怎样形成的很重要，那是下一章的主题。

19. 半衰期是一个样本中有半数粒子衰变掉的时间。如果开始时是 100 个中子，那么 13 分钟后就剩 50 个，26 分钟后剩 25 个，39 分钟后还有一打，依此类推。这是一种统计行为，其中单个粒子的命运是不可预测的，也就是无法预测原有中子中的哪一个将在头 13 分钟内或在任何一个后继期间内衰变，但粒子数当然必定远多于 100，而大群粒子的整体行为是完全可以预测的。这是量子物理学最引人注目的特征之一，它也适用于所有放射性

衰变及其他过程，并不只是中子衰变。

20．那么这浓汤又从何而来呢？在20世纪40年代（甚至在50年代、60年代和70年代），这个问题是无法回答的。伽莫夫绕开了这个难题，办法是援引某一个弗里德曼模型，即宇宙从稀薄状态开始，收缩成稠厚的"汤"，然后再膨胀。但是现在有了走出困境的其他方案，那是本书第三篇的主题。

21．那个把铁56放在谷底的光滑峡谷当然是过于简化了。要进一步比拟，氦4核就是在峡谷一侧的一个坑里，把它赶出来可不容易；而含8个粒子的核就可以想象成是在峡谷一侧一个很尖的小山包的顶上，稍受扰动就跌落下来。

22．*The Creation of the Universe*，p.65.

23．开氏温标中的零度，若取最接近的整数是－273℃。所有粒子在绝对零度都处于最低能级，称作零点态，没有任何东西能有比这更低的温度。

24．André Deutsch ed.，pp.131—132.

第五章　宇宙的两把钥匙

20 世纪 30 年代的天体物理学家对太阳和恒星的兴趣比对宇宙起源要大得多。创世的奥秘似乎仍然更多地是形而上学家沉思的课题，而较少是科学研究范围内的事情；恒星怎么发光的问题倒是既很有吸引力，又正好能运用 20 世纪 20 年代物理学的革命性新发现来探索。但结果却是，对恒星如何保持高温的研究直接导致了对宇宙整体和大爆炸本身的更好认识。

太阳是离得最近、我们也知道得最多的恒星。如果天文学家能有什么希望认识恒星的普遍活动方式，他们就必须先对太阳有至少是大略的了解。但是在开始时，似乎连新物理学即量子物理学都不足以胜任此项任务。

所有关于地球已存在了 40 亿年到 50 亿年的证据都清楚地表明太阳也应该有相似的年龄。19 世纪的物理学家已经认识到通常的化学燃烧过程不可能使太阳在这么长时间内维持高温。例如，让一个像太阳那么大的煤球在纯氧中燃烧，并且每秒钟产生的热量与太阳现在所

释放的相同，那么这个煤球在大约 1500 年里就会烧光。19 世纪后半叶，德国物理学家亥姆霍兹(Hermann Helmholtz)和英国物理学家威廉·汤姆孙(William Thomson，即后来的开尔文勋爵，绝对温标就是为纪念他而取名的)首次试图运用天体物理学过程，即在恒星里而不是在地球上发生的过程，来解释恒星如何能维持高温如此之久。

亥姆霍兹和汤姆孙都是当时科学界的大人物，兴趣范围都很广，而地球的年龄乃至太阳的年龄是他们共同关注的问题。亥姆霍兹于 1854 年得出该年龄是 25 000 000 年，汤姆孙稍后得出的要长些，他认为最可能值是 1 亿年。现在知道即使汤姆孙的估计也小了好几十倍，但这些结果已是从当时一些教堂仍认可的世界创始于几千年前的观念走出来的很有意义的一步(达尔文的《物种起源》是 1859 年发表的；亥姆霍兹和汤姆孙所作的地球年龄估计远不是什么小圈子里的科学资料，而是与当时那场最重大的科学和哲学论争直接相关)。问题是，即使太阳只保持高温 1 亿年，其能量又从何而来呢?

答案似乎是“引力”。如果太阳开始时是空间中的一团稀薄气体云，然后在自身引力作用下收缩成较密集的球，那么它就会随着收缩而升温。当你拉伸一根弹簧时，你必须做功(加进能量)以克服弹簧的弹力；当你松开弹簧时，能量就释放出来。当你把一个重物从地面上举起时也是同样情形。你把引力势能形式的能量加给了重物；当你松手时，能量随重物的降落而转变成动能，当重物的降落被地面阻住时动能又转变成热能。所有组成太阳的粒子都“想要”落向中心，落向由它们组成的系统的质量中心。如果它们果真降落，它们就将释放出引力势能并最终转变成热，就像下落的重物以热的形式放出能量一样。一颗较致密的恒星比一颗较疏松的恒星处在更低的能量状态，因为前者的组成粒子更靠近质量中心。[1]所以如果开始时有一颗与太阳

相似但稍大一些的恒星，让它在自身引力作用下缓慢收缩，就可以预期热会产生出来。

太阳的年龄

天文学家能从太阳系中天体的轨道运动、地球上潮汐的强度及其他途径很好地计算出太阳的距离(即天文单位)和质量。他们知道太阳每秒钟辐射多少能量，因为他们知道它必须实际上有多亮才能在天空中如看上去的那么亮。它释放的能量是每秒 4×10^{33} 尔格(4×10^{26} 焦)，即每年约 10^{41} 尔格(10^{34} 焦)。当亥姆霍兹和汤姆孙计算一颗像太阳这样的恒星若缓慢地收缩将释放多少能量时，他们得出从稀疏气体云到恒星的收缩所提供的能量能使恒星维持上述强度的辐射 1000 万年到 1 亿年之间，然后恒星内部就会冷下来并发生另一种猛烈的收缩。这个时间尺度现在称作开尔文—亥姆霍兹时标，这无论是从姓氏字母顺序或是作计算的时间先后来说对亥姆霍兹都有点不公正。但这种收缩在恒星生命的早期无疑是重要的，正是这个过程的引力加热作用使恒星升温到开始发光。但到 20 世纪 20 年代已经清楚，地球和太阳肯定(因而恒星也应该)比开尔文和亥姆霍兹所想的要古老得多，即不是几千万年而是几十亿年。那么能量究竟从何而来以维持每秒 4×10^{33} 尔格(4×10^{26} 焦)的辐射如此之久，使太阳今天依然光芒万丈？[2]

开尔文对地球年龄的估算是基于地面温度与很深的矿井下温度的比较。热在从地球内部泄露出来，通过估计这种热散失的速率并沿时间反推，他推断我们的行星是在几千万年前融合成的。开尔文和天文学家们都认为这个地球年龄与开尔文—亥姆霍兹太阳年龄的相符是有说服力的。但即使在 19 世纪后期，地质学家们已经有清楚的证据表明地球年龄必定远大于 10 000 000 年，否则就没有时间形成像阿尔卑

斯山的那种很厚的褶层岩石结构，而且进化论生物学家也宁愿地球有更长时间的历史，演化进程才能产生出今天地球上复杂的生命形式。开尔文所估计的地球年龄是错的，其原因在世纪之交由于放射性现象的发现而变得清楚。放射性元素天然存在于世界各地的普通岩石中，放射性活动的机理是原子核分裂成两个或更多个部分并射出 α 粒子或别的粒子而转移到较低的能态。放射性核衰变时的能量改变也表现为热，流过地壳岩石的热中大约有 90%来自放射性活动。所以开尔文的地球年龄必须扩大 10 倍到 100 倍，才能对太阳和行星已经存在了多久给出一个更精确但仍嫌粗糙的指引。

20 世纪 20 年代有几项进展使得对太阳和其他恒星的内部活动开始有了较好的认识。第一，光谱分析显示出太阳物质中约 70%是氢，28%是氦，仅 2%是重元素。然后爱丁顿几乎一手创立了天体物理学这门学科，并从对双星的研究中发现恒星的光度与其质量直接相关。这一于 1924 年宣布的发现是认识恒星活动的关键一步，1926 年爱丁顿的里程碑式的著作《恒星的内部构造》出版，其中论述了如何能用描述一个巨大热气体球行为的基本物理定律来解释恒星的这一性质和其他性质。与此同时，继卢瑟福的先驱性工作之后，对原子的探测和对放射性的研究，使物理学家开始熟悉了(或至少是不再那么陌生了)1905 年就已发表的爱因斯坦狭义相对论的一个更惊人的预言。

这就是质量与能量的等价和可以相互转变，也就是 $E=mc^2$。随着物理学家能够更精确地测量原子(从而测量原子核)的质量，随着他们对使原子核保持在一起的力的认识的缓慢进展，终于清楚当一个重核分裂成两块较轻的碎片时，所释放的能量是精确地由碎片质量的损失来补偿的。当核 A 分裂成 B 和 C 时，B 和 C 质量之和小于 A 的质量，差额正好是按 $E=mc^2$ 的定则提供了裂变释放的能量。在 20 世

纪 20 年代，更多的是在 30 年代，用新的量子理论同时也用相对论武装起来的物理学家，开始理解了爱因斯坦理论对核世界的意义。

使能量之得与质量之失相当的规则同样可以应用于处在稳定性峡谷另一侧的核。当然在这种情况下是两个较轻的核合在一起形成一个较重的核，这个过程也有质量损失和相应的能量释放。例如，一个氦 4 核由两个质子和两个中子组成。物理学家以碳 12 原子的质量来计量原子和核的质量，该原子含有 6 个质子和 6 个中子(加上核外 6 个电子的云)，其质量被定义为 12 个原子质量单位。按照这种单位，质子质量是 1.007 275，中子质量是 1.008 664。把两个质子和两个中子加在一起“应该”得到总质量 4.031 878。但是 α 粒子即氦核的质量是 4.001 40 个原子质量单位。稍多于 0.03 个单位的质量“损失”了，并且每次都在那 4 个粒子结合成 1 个氦核时作为热能出现。转变成能量的质量大约是初始总质量的0.75%，由于这个质量要乘以光速的平方，而光速是 3×10^{10} 厘米/秒，结果就非常可观，但前提是要找到在恒星内合成氦的途径。[3]

在 20 世纪 20 年代后期和 30 年代初期，爱丁顿和他的同行们并没有得出一个关于 4 个氢核怎样转变成 1 个氦核的详细理论。但他们知道在这个过程中质量将损失并转变成能量，也知道太阳每秒钟辐射多少能量。运用爱因斯坦公式就很容易算出，太阳要维持现在的亮度就得每秒钟把 4 000 000 吨的物质转变成能量，也就是说每秒钟要有稍多于 6 亿吨的氢(质子)转变成氦(α 粒子)。这听起来很多，但其实是太阳质量的很小的一部分。太阳质量若取整数是 2×10^{33} 克，或 2×10^{27} 吨。太阳每秒钟把 6 亿(6×10^{8})吨氢转变成氦，也就是每年“烧掉”不到 2×10^{16} 吨的燃料。它在 1000 年里烧掉 2×10^{19} 吨氢，在 1 000 000 年烧掉 2×10^{22} 吨。即使以这种速率烧 100 亿年也只用掉 $2\times$

10^{26}吨氢，即太阳总质量的10%。这些燃料的0.7%即1.4×10^{24}吨的物质在这个过程中全都转变成了能量。在氢以这种速率“烧”成氦的过程进行了约100亿年后，太阳的物质成分才改变到使其外观发生重要变化。这样，地球年龄和太阳年龄立即再次调和了。如果地球已是40亿或50亿岁，而且太阳这样的恒星稳定地燃烧氢能达100亿年之久，那么太阳处在我们现在所知这种状态的时间就正好大约过了其全部时间之半。

尽管这个理论的大致轮廓看上去的确很好，当爱丁顿提出氢聚变成氦的过程必定是太阳和恒星的能量源泉时，与他同时的物理学家中却有许多不予理睬。这在当时是有缘故的。爱丁顿将基本物理定律应用于太阳和恒星结构时包括了这样一项计算，即太阳中心必须保持多高温度，才能提供它那可见的表面光亮，并产生足够强的压力来对抗向内的引力以维持它的稳定。该温度大约是15 000 000开。温度就是为大量粒子均分的动能的量度，我研究的空气温度就表征了空气中一个典型分子平均说来运动得有多快。在太阳中心，温度就是一个典型质子与其近邻反复碰撞弹来弹去的运动平均来说有多快的量度。温度越高，质子的速度也越大，它们相互之间的和施于恒星上层物质的压强也越大，这当然也就是爱丁顿的温度计算的思路。问题是，他得到的15 000 000开的温度似乎并未高到足以使聚变过程发生。

两个质子（且取最简单的情况）的聚变只有在它们有足够高的能量克服电斥力而碰撞时才会发生。回到那个火山比拟，两个粒子必须具有足够大的速度从山峰的相反两侧冲上山坡、越过峰顶，才能落到火山口内，即进入强核力支配的范围。一个温度为15 000 000开的质子海，即使有着太阳核心的密度，也并不具有足够高的能量

使次数足够多的碰撞发生从而为太阳提供热量，当物理学家起初计算时的情况看来就是这样。要么是那些物理学家错了，要么是爱丁顿的计算错了。

爱丁顿对自己的那些数字很有信心，据广泛报道他曾告诉那些抱怀疑态度的同事“往前找更热的地方”（换句话说是往地狱去）。难题很快就将解决，这在很大程度上要归功于伽莫夫得出的描述 α 粒子怎样从大质量核的势阱（火山坑）中出来的方程。改变对聚变过程的认识的关键是一种称为不确定性的量子物理现象，这在下文中将会更详细地叙述。简括地说它的意思就是，如果两个质子靠拢发生碰撞，它们并不是必须要越过由电力造成的势垒。只要它们是在垒的顶峰附近，就有可能沿一条隧道穿过去而落到势阱内，哪怕它们并不十分具有足够的能量来翻越峰顶。隧道效应提供了即使在15 000 000开的温度也能有效地发生聚变的途径。

于是在 20 世纪 30 年代中期理论站住了脚。至少对专家来说已经清楚，太阳能量的来源必定是核聚变及相随的质量到能量的转变。由于太阳是由 70%的氢和 28%的氦组成，又由于氦核特别稳定，看来关键过程必定是每 4 个氢核（即 4 个质子）转变成一个氦核也就很清楚了。但是没有人确切地知道质子是怎样转变成氦核的，直到 1939 年才由贝特（后来以 αβγ 论文的“缺席”中间作者而名垂不朽的那位）找到了一条巧妙途径。他的工作开创了天体物理学的一个新领域，即研究怎样在恒星内部制造元素（恒星核合成）。这个领域是（在第二次世界大战造成的短暂延误之后）由新的一代天体物理学家承担起来的，并且导致诺贝尔奖于 1967 年授予贝特本人，又于 1983 年授予福勒。

恒星中的循环和链条

贝特1906年出生于斯特拉斯堡(当时属德国，现在是法国的一部分)，在法兰克福大学和慕尼黑大学学习，1928年获博士学位后在德国的大学里教物理学和做研究直到1933年。由于纳粹在德国的兴起，他去了英国，在曼彻斯特大学待了一年后又去了美国的康奈尔大学，并加入了美国国籍。20世纪40年代贝特参与了曼哈顿计划即第一颗原子弹的设计和制造，是该计划的理论物理学家组组长；后来又作为代表参加了日内瓦的第一次国际禁止核试验会议，为在核大国间达成禁止大气层核武器试验的协议出了力，在相应的条约签署之后仍继续做核裁军问题上的顾问。他一生中获得了许多荣誉。他的主要科学贡献则是于1938年作出、1939年发表的工作，其中论证了在恒星内部怎样通过制造氦来产生能量。主要是这项工作使他获得了诺贝尔奖。[4]

贝特提出的恒星内部产生能量的第一个机制，除氢以外还需要重元素尤其是碳的存在。他计算出，在适当条件下一个质子与一个碳12核的碰撞能生成一个氮13的核，然后又放出一个正电子变成碳13核。碳13又与一个质子(氢核)碰撞生成氮14，再与一个质子碰撞生成氧15，然后又放出一个正电子变成氮15。当又一个质子加进来时，最可能的结果不是生成一个氧16核，而是4个核子作为一个α粒子分裂出来，形成一个氦核并留下一个碳12核，正好又回到了起始条件。在大约1000次氮15与质子的碰撞中有一次会生成氧16，即使如此，当又有2个质子先后加进来时，就会放出一个α粒子而衰变成氮14，也加入了循环。净效果是4个质子转变成了1个氦核，并释放出相应的能量。由于碳12本身并不消耗，而只是作为催化剂来使聚变循环不断进行，所以只需要很少一点碳就能进行大量的核聚变

并产生大量能量。

这个过程除氢和氦外还有碳、氮、氧核加入，循环进行，起点和终点都是碳 12，把 4 个质子转变成 1 个氦核，于是就很自然地称作碳氮氧循环或 CNO 循环。[5]贝特和他的一位同事克利奇菲尔德(Charles Critchfield)后来又提出了恒星中氢聚变的另一种可能途径。这是一个逐步过程，从氢核开始，先造出氘，然后是氦 3 和氦 4；这两种氦核又可以碰撞生成有 7 个核子的核，再加上 1 个质子就转变成两个 α 粒子。这个过程在上一章中讲过，它从两个质子碰撞生成氘开始，而以氦 4 核告终，称为质子—质子(或 pp)链。

贝特的计算表明，这些反应可以在爱丁顿已经证明恒星内部必定具备的温度和压强条件下进行。太阳中主要的产能过程被认为是 pp 链，它以所需要的效率在约 15 000 000 开的温度下运行； 而 CNO 循环则更适于在更高的温度即20 000 000开以上运行，因而在更大质量的恒星中更重要，因为这些恒星内部的温度就会要更高才能抵抗住引力收缩的可能性。不过这些有趣的细节并不与探索大爆炸的故事直接关联。

表 5.1　CNO 循环

$C^{12} + H^{1} \longrightarrow N^{13} + \gamma$

$N^{13} \longrightarrow C^{13} + \beta^{+} + \nu$

$C^{13} + H^{1} \longrightarrow N^{14} + \gamma$

$N^{14} + H^{1} \longrightarrow O^{15} + \gamma$

$O^{15} \longrightarrow N^{15} + \beta^{+} + \nu$

$N^{15} + H^{1} \longrightarrow C^{12} + He^{4}$

或，1000 次中有 1 次：

$N^{15} + H^{1} \longrightarrow O^{16} + \gamma$

$O^{16} + H^{1} \longrightarrow F^{17} + \gamma$

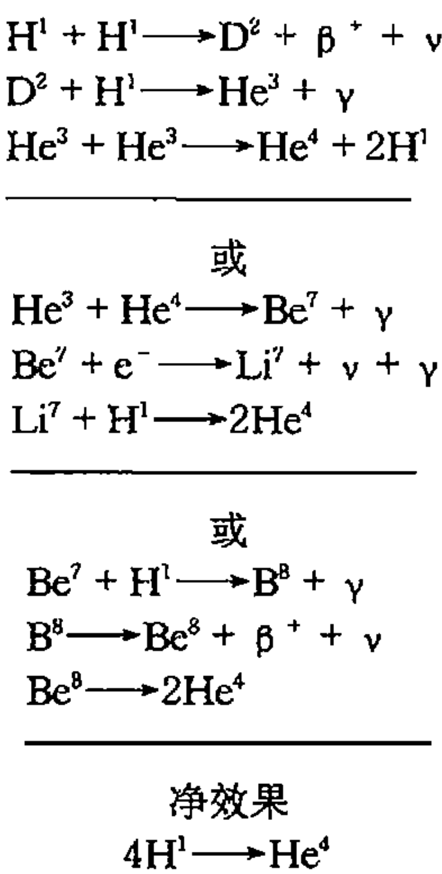
$F^{17} \longrightarrow O^{17} + \beta^{+} + \nu$

$O^{17} + H^{1} \longrightarrow N^{14} + He^{4}$

净效果

$4H^{1} \longrightarrow He^{4}$

表 5.2　pp 链



$H^1 + H^1 \longrightarrow D^2 + \beta^+ + \nu$
$D^2 + H^1 \longrightarrow He^3 + \gamma$
$He^3 + He^3 \longrightarrow He^4 + 2H^1$

或

$He^3 + He^4 \longrightarrow Be^7 + \gamma$
$Be^7 + e^- \longrightarrow Li^7 + \nu + \gamma$
$Li^7 + H^1 \longrightarrow 2He^4$

或

$Be^7 + H^1 \longrightarrow B^8 + \gamma$
$B^8 \longrightarrow Be^8 + \beta^+ + \nu$
$Be^8 \longrightarrow 2He^4$

净效果

$4H^1 \longrightarrow He^4$

烹制元素

贝特的工作并没有解释恒星中的碳起初是怎么来的，但的确解释了太阳这样的恒星怎样从氢到氦的转变过程中获得能量。他的计算所依据的是在实验室条件下对粒子相互作用的能力即截面的测量，然后测得的这些截面又必须外推到按照物理定律在恒星内部必定存在的条件下去。这在科学上是巨大的一步，不仅在于表明了实验室里测量的粒子行为能够外推而揭示出恒星如何运转，而更在于改变了天体物理学的整个观念。由于爱丁顿和贝特，天体物理学成了一门实验科学，地球上的粒子束碰撞实验现在可以揭开恒星核心核聚变的秘密。当贝特的第一篇关于 CN 循环（当时是）的论文于 1939 年出现在《物理学评论》上时，[6]在加州理工学院凯洛格辐射实验室工作的一群核物理学家感受到了这样一种冲击。

在那个实验室里，资深物理学家查尔斯 · 劳里森（Charles Lauritsen）和两个年轻人，一个是他儿子托马斯 · 劳里森（Thomas Lauritsen），另一个就是福勒，正在测量碳核和氮核受质子束轰击的

作用截面。贝特的论文使他们知道自己是在实验室里研究太阳和恒星中发生着的过程。44 年后，福勒在他的诺贝尔奖演讲[7]中说“它给了我们持久的印象”。这个印象的确很深，以致当 1946 年实验室恢复对原子核的基础研究工作时，老劳里森决定集中力量研究那些据信在恒星中发生的核反应。福勒接受了任务，并成为该项研究的领头人。

福勒是一位热情奔放的外向型人物，今天依然在学术上非常活跃*。他在自己被授予诺贝尔奖的决定宣布时曾对记者说：“我想要保持活跃，直到我被送走为止。”他 1911 年生于匹兹堡，在俄亥俄州立大学学物理，1933 年毕业，然后去了加利福尼亚，1936 年在加州理工学院获博士学位。从此以后他的主要根据地就在那里，尽管也时常在世界各地的其他研究中心待上几个月。福勒和凯洛格实验室对认识恒星核合成和对精确计算大爆炸生成的氦的数量都起了关键作用。但是正如福勒在他的诺贝尔奖演讲中承认的，恒星核合成的“主要概念”是来自霍伊尔(现在的霍伊尔爵士)**于 1946 年和 1954 年发表的两篇论文。

这两个时间是重要的，尤其是对霍伊尔来说。当宣布福勒被授予诺贝尔奖时，许多有关报道都提到了霍伊尔的贡献，其中有些报道很快还指出，找出在恒星中合成元素的途径是对这位稳恒态理论创立者之一的明显需求。要知道，如果没有大爆炸，稳恒态宇宙就必须在恒星中造出元素。由于稳恒态理论已经失败，那些报道便说，霍伊尔是以错误的动机得到了正确的结果，并且暗示这或许就是他没能与福勒共享诺贝尔奖的缘故。但是如霍伊尔在福勒获奖时对我指出的，稳恒态理论直到 1948 年才被提出，而他关于核合成的第一篇论文是写于

* 福勒已于 1995 年逝世。——译者

** 霍伊尔已于 2001 年逝世。——译者

1945 年、发表于 1946 年。[8]除此之外，大爆炸理论家的确无论如何也不能在大爆炸中造出任何比氦更重的东西。

霍伊尔给人的外表印象是易于动怒，不能容忍庸人。这使得他不为自己的工作机构所喜爱，并于 1973 年 58 岁时就过早地从剑桥大学的职位上退休。但实际上他羞怯、内向，竭尽全力于工作，又显然善于把自己的重要思想传达给别人。他 1915 年出生于约克郡的宾利，有时他的反对者说他因此养成了“约克郡的生硬言谈”习惯，实际上有些似乎粗鲁的表现应该被视为北方人传统的直率而予以谅解。但是，英国的学术机构没有能充分认识他的研究和他关于科学应该怎样发展的思想的功绩，这显然使他深受伤害。假如他真如有些人想象的那样厚脸皮和不敏感，他将肯定会在剑桥任职更久。

不过，他的早期生涯也是沿着常规路线，从家乡的中学到剑桥的伊曼纽尔学院，然后于 1945 年成了圣约翰学院的研究员。他于 1958 年成为普鲁密安教授，并且是 1967 年建立的剑桥理论天文学研究所的鼓动者和首任所长，但这是他在学术机构职务的顶峰。尽管他还在许多高级委员会任职(同时还能有时间写科学普及和科学幻想著作)，并被选为皇家学会会员和被授予爵士称号，他却在关于剑桥大学天文学研究的地位和如何发展的问题上与校方闹翻了，还在应如何资助和发展整个科学研究的问题上与英国科学管理官员持强烈不同的意见，并且在 20 世纪 70 年代和 80 年代由于其普遍被认为荒谬的关于生命起源的主张而疏远了许多同事。但是所有这些，或是天文学家们对稳恒态模型和他关于宇宙本质的其他任何观点所作的任何评价，都不可能贬低他对发现恒星内部如何烹制元素所起的关键作用。

元素生成问题使天文学家开始感到紧迫，并不是因为他们对宇宙的大爆炸或稳恒态模型的什么兴趣，而是因为他们在 20 世纪 40 年代

和50年代不断改进的恒星光谱观测日益清楚地表明，不同的恒星含有不同数量的不同元素。可以设想恒星形成的材料是来自大爆炸，也可以设想那些材料是在星系之间的空间不断地被创造出来。但既然发现恒星的成分有系统差异，有的恒星重元素含量比其他恒星多，那就得开始考虑有些重元素是在恒星自身内由原初材料（不管它们可能是什么）制造出来的。

霍伊尔1946年的文章首次在公认的恒星结构和演化的框架内，采用当时有关核反应速率和截面等等的最佳资料，对核合成的基本思想作了清楚的说明。就在伽莫夫小组为找到在大爆炸中生成比氦更重的元素的方式而奋斗时，霍伊尔也在努力寻求在恒星内合成重元素的途径，他于1953年首次访问了加州理工学院，并很快就与福勒通力合作。关键问题是如何越过不稳定的硼8核。唯一的途径是仰仗一种三体碰撞，即三个 α 粒子几乎同时碰到一起形成一个碳12核。伽莫夫不可能让大爆炸来这样做，因为即使在创世瞬间的几分钟后，宇宙中的物质已铺散得太稀薄，温度也太低，不足以使这种碰撞多到能生成今天宇宙中的重元素。但在恒星内部就既高温又密集，并且这种条件能保持成百上千万年之久，于是就提供了好得多的机会使相对罕见的三体碰撞能多到足以产生所需数量的碳。

想法看来挺好，但也遇到了问题，与爱丁顿在物理学家告诉他太阳温度不足以使氢聚变发生时所遇到的问题有点相似。一位物理学家萨尔皮特（Ed Salpeter）在1951年访问凯洛格实验室时作了计算，发现截面仍不够大，固然可以在恒星内生成一些碳12，但数量不够。就在这时霍伊尔作出了戏剧性的贡献。他在1953年来到加州理工学院时已经相信所有重元素都是在恒星内生成。他从另一头出发作了计算，由恒星中重元素丰度的观测结果（来自光谱分析）来推求三体碰撞

反应必须进行得有多快，结果是必定要比萨尔皮特计算的快得多。于是他预言碳 12 必定能以一种所谓激发态存在，即有着高于最低能态的能量。只有存在这种有着恰当能量的激发态时，三个 α 粒子的碰撞才足以生成在恒星光谱中观测到的所有的碳。反应是由一种在三个 α 粒子的能态与碳 12 核的能态之间的所谓共振过程来促进的，而共振只有在碳 12 处于恰当能级时才会发生，这也就是霍伊尔何以能预言碳 12 必定在何种激发能态的缘由。

霍伊尔于是去纠缠加州理工学院的物理学家，直到其中一组人开始去寻找碳 12 的激发态为止。他们所采用的反应是氘粒子与氮 14 核碰撞生成碳 12 和一个 α 粒子，结果与霍伊尔所预言的几乎完全一致。

但这离证明碳 12 激发态能由三个 α 粒子的反应生成还差一步，现在轮到福勒和劳里森父子还有库克(Charles Cook)来由硼 12 的衰变制造出激发的碳 12 了。他们发现，尽管有的激发态碳 12 能落回到最低能态(即基态)并保持为碳 12，另一些却又分裂为三个 α 粒子。反应是可逆的，既然激发态碳 12 能衰变成三个 α 粒子，无疑三个 α 粒子也能结合成激发态碳 12。这就证明了如同燃烧氢生成氦一样，恒星也能燃烧氦来生成碳。氦燃烧过程解释了被称为红巨星的大个子恒星如何保持高温，并且也使天体物理学家越过了核合成进程在第 8 号元素处的障碍。当然，它还提供了使 CNO 循环运转所需要的碳。通过查看恒星光谱学资料，霍伊尔正确地预言了物理学家在地球实验室里将会得到的结果。这又使他们有了信心去继续在实验室里测量反应速率，并利用所得结果去计算出生成天然存在于恒星中的所有元素的所有同位素的整个反应链。

作为对恒星核合成认识进程的简略评述，碳 12 以后元素的生成

就都可以一笔带过了，正如在霍伊尔和福勒及其同事已经做好的蛋糕上加点糖。概而言之，现在已经容易理解所有的元素如何在恒星里造出来。将 α 粒子加进核里，每次都使核的质量增加 4 个单位，核又会衰变射出电子、正电子或中子，成为其他元素和同位素。对很重（比铁更重）的元素来说，捕获单个中子使核质量每次增加一个单位也是重要的。再也没有像从 α 粒子造出碳 12 那样使天体物理学家苦恼的大裂口。剩下的事只是辛苦地、仔细地找出所有需要的截面和反应速率，使之与恒星内的温度和压强条件并与由恒星光谱观测得到的元素丰度相符合。

截面测量并非易事；把测量结果从实验室加速器里碰撞的相对较高能量条件外延到恒星内粒子间碰撞的低得多的能量条件，更需要很高的技巧；而观测家们也得很紧张地查明宇宙中物质的成分。但所有这些都在 20 世纪 50 年代中期完成了。福勒在剑桥度过了 1954—1955 学年，与霍伊尔还有玛格丽特 · 伯比奇（Margaret Burbidge）和杰弗里 · 伯比奇（Geoffrey Burbidge）一道工作，后两位是英国的一对天文学家夫妇。在福勒返回加州理工学院后合作又继续了很久，霍伊尔或伯比奇夫妇或三人一起也去过凯洛格实验室。1956 年，天文学家聚斯（Hans Suess）和尤里（Harold Urey）发表了所有天然存在元素相对丰度的最好资料；那 4 位合作者则于同年在《科学》上刊登了一篇关于元素起源的短文，接着就是登在《现代物理评论》上的那篇至今被视为科学经典的论文。[9]论文作者按姓氏字母顺序是伯比奇、伯比奇、福勒、霍伊尔，所有天文学家都知道该文被简称为 B^2FH，引用时无需多作说明。文章描述了除氢和氦以外所有天然存在的核是如何在恒星内生成的，用瑞典科学院宣布福勒获诺贝尔奖时的话，该文“仍然是我们对这个领域知识的基础，核物理学和空间研究的最新进展已经

进一步证实了其正确性”。我的一本个人笔记使我回想起1966年作为研究生第一次读那篇文章时的激动和敬畏之情，因为我知道那些方程式解释了我体内除原初的氢以外的所有原子是从哪里来的，这些原子又怎样在恒星中被烹制出来。如同任何一项科学研究最多所能企及的，这篇文章为一个研究篇章划上了句号，对一个不只是科学上而且是哲学上的重大难题作了完整解答。从霍伊尔发表他关于这个课题的第一篇标志性论文算起，他和合作者为这个解答奋斗了整整10年。

除了一些细节有待进一步搞清楚外，B^2FH论文标志着恒星核合成难题的终结。但是它并不标志着所有核合成难题的终结。从宇宙中第一代恒星中的70%到75%的氢和25%到30%的氦出发，天体物理学家能够说明所有的重元素是怎样合成的，还能猜测这些重元素是在一些年老恒星作为新星或超新星爆发时散布到空中，并加入到形成新恒星的气体之中。我们的太阳是相对年轻的，它含有进入再循环的很久以前死亡的老恒星里的物质，这就是它的2%重元素以及组成地球和我们人体的材料之由来。人体重量的65%是氧，18%是碳。这都来自恒星内的3α粒子捕获过程及其他反应。但是恒星核合成不能解释为什么宇宙中有这么多氦。20世纪50年代后期和60年代初期的天体物理学家为这个“氦问题”而苦恼，尽管有此前伽莫夫和阿尔弗在大爆炸中造出氦的成功。也许是因为他们原来的由大爆炸制造出所有元素的意图的失败，他们对氦的成功也被忽略了；无论原因是什么，把改进的核反应速率知识运用到被认为在宇宙创生后的头几分钟里应有的条件中去的任务，又留给了福勒和他的学生瓦戈纳(Robert Wagoner)，还有再次出场的霍伊尔。[10]

福勒小组在计算中包括了由他们自己测定的将近100个核过程的反应速率，并且用实验证明了其他反应都不重要，从而确认了大爆炸

不可能产生出足够数量的任何一种比氦重的元素，而且由大爆炸生成的氦4的比率应该是25%，相随的还有氘、氦3和微量的锂7，这几种元素的比率都与在太阳系里观测到的类似。

瓦戈纳、福勒和霍伊尔在1967年发表了他们的结果。这里又有我个人的一次重要经历，我第一次去剑桥就是去听该小组关于那些结果的报告，我现在能清楚地回忆起当时不为人知的研究生霍金在会场上所提的尖锐问题。那次访问，和会议带给我的兴奋，促使我从萨塞克斯大学转到了剑桥，并打算做宇宙学研究。虽然我后来是以关于恒星结构的工作而告终，但我仍然认为那次旅行很有价值。当然更重要得多的是，瓦戈纳、福勒和霍伊尔的成果给了宇宙学一个重大冲击。正如麦克雷爵士所说："正是这篇文章使许多物理学家承认热大爆炸宇宙学是一门严肃的定量的科学。"[11]

科学的确是"认真"的，还能提出一个比宇宙中的氦丰度精细得多的问题。氦是由氘核即重氢核的聚变生成的。几乎全部的氘核都这样用掉了，但是根据对恒星和星系的光谱研究，现在宇宙中氢的总量里看来有很小一部分，即0.01%到0.001%之间，是以氘的形式存在。虽然由大爆炸产生的氦的比率并不很敏感地与大爆炸产生的物质总密度相关，氘丰度却是物质密度的一个很敏感的指示器。计算中用的模型的密度越大，反应就进行得越快，氘核就消耗得越快。瓦戈纳、福勒和霍伊尔所作的计算表明，宇宙中常见物质的密度小于使宇宙闭合并有朝一日坍缩回到火球所需要的临界值。这是有利于宇宙开放并将永远膨胀下去这种可能性的最强证据。但这并不是最后结论。本书第三篇将要介绍的宇宙学最新认识中就有宇宙里存在着大量其他物质这种可能性，那些物质并不参与瓦戈纳、福勒、霍伊尔及他们的后继者所描述的核反应，这就有可能使宇宙由引力来"闭合"，而无

论由氘所指示的来自大爆炸的普通物质(我指的是原子和原子核)是多少。但对20世纪60年代后期来说，这是后来的事情。

核合成研究到那时已经提供了一对吻合得极好的证据支持热大爆炸模型。自伽莫夫以来的所有研究都表明，任何比氦重的元素都不可能在大爆炸中生成。那些元素就必须以别的途径制造出来。恒星核合成的反应式表明，所有比氦重的元素的确能在恒星中生成，但在宇宙中观测到的氦却不可能是由此而来，它们必须来自别处。大爆炸理论需要恒星核合成，恒星核合成也需要大爆炸。热大爆炸与恒星内的核合成结合在一起，描绘出了一幅一切从何而来的完美图画。

在那篇1967年的论文中还有另一段“定量科学”使物理学家对大爆炸理论突然关切起来。瓦戈纳、福勒和霍伊尔首次把对宇宙微波背景辐射的新发现定量地运用于确定他们的模型参量。这个宇宙背景就是伽莫夫及其同事在20世纪50年代部分地预言过但后来被遗忘的辐射。这是宇宙的确从一个火球中产生出来的第二个强有力的证据，也是宇宙的第二把钥匙。

失落的年代

测量宇宙的温度并从中找出有关宇宙从中诞生的大爆炸的更多信息的思想，是太难以要求20世纪50年代的物理学家和天文学家去认真看待了。但这并不意味着这个思想被完全忽视，不只一个天文学家颇为悔恨地回顾过那个10年，并自责没有能把这个思想推进到其合乎逻辑的结论。的确，伽莫夫、阿尔弗和赫尔曼必定有这样的感慨，因为一些清楚地提示宇宙的大约3K背景温度的天文观测在20世纪30年代就已经作过，而在50年代伽莫夫和他的同事肯定是知道这些观测的。

这些观测与其他许多来自空中的信息一样，也依赖于光谱学。20世纪30年代的天文学家已首次识别出与星际空间分子对应的光谱特征。恒星光携带着存在于恒星大气中的原子(更准确地说是离子，即失去了一些电子的原子)的光谱印记。那些电磁谱中的特征线是明亮的发射线(辐射能量)或暗黑的吸收线(吸收来自恒星下层的能量)。那些谱线的强度和由谱线显示出的电离程度使天文学家得以推断遥远恒星的温度并确定其化学成分。但是有些光谱中的有些谱线却对应着在恒星表面温度下不可能稳定存在的化合物。其中最早证认出的一个是氰即CN，由一个碳原子和一个氮原子稳定配对组成的原子团。这种化合物不可能在恒星中存在，因为那里的高温会使它们很快分解成原子，但却可以存在于恒星之间的冷暗气体和尘埃云中。当远处恒星的光穿过这些云时，那些化合物就会在恒星光谱中加进暗线，它们的存在就是这样揭示出来的。

正如恒星光谱能展示恒星的温度一样，那些吸收谱也能展示星际物质云的温度。1940年，威尔逊山的亚当斯(W. S. Adams)观测到了一种较高能态氰的星际谱线，加拿大自治领天体物理台的麦凯勒(Andrew McKellar)从中得出星际云的温度约为2.3K。到1950年这个结果已写入光谱学的标准教科书中[12]并广为天文学家所知，包括伽莫夫在内。但是没有人想过把空中这些寒冷物质云的温度理解为“宇宙的温度”。最堪称擦肩而过的一次失误是在1956年，当霍伊尔和伽莫夫坐在一辆崭新的白色凯迪拉克牌有篷汽车里，在南加利福尼亚的拉霍亚一带行驶的时候。

霍伊尔在1981年出版的《新科学家》杂志的一篇文章里细述了这个故事。1956年夏天他去见福勒和加州理工学院的其他同行，伽莫夫从拉霍亚打电话，邀请福勒、霍伊尔，还有伯比奇夫妇去作客。

伽莫夫那时在拉霍亚作两个月的一般动力学顾问，那份差使很赚钱（两个月顾问费买了那辆白色凯迪拉克小汽车），而且显然几乎不用真干什么事，但他必须待在拉霍亚的什么地方（哪怕是在海滩上）以备不时之需。B^2FH 队伍不算太勉强地向南进发了。当时伽莫夫对现在宇宙温度的估算是在 5 开到几十开之间，而霍伊尔作为稳恒态派则认为根本就没有什么背景辐射。真理这次真是从他们两位的鼻子下滑了过去。霍伊尔写道：

> 有时乔治和我想出去，就两个人谈谈。我记得乔治开着那辆白色凯迪拉克带着我兜风，他向我解释他为什么确信宇宙必定有一个微波背景，我告诉他宇宙不可能有他所主张的那么高温度的微波背景，因为麦凯勒已经由对 CH 和 CN 原子团的观测为任何此类辐射的温度给定了一个 3K 的上限。也许是凯迪拉克车里太舒服了，也许是乔治要一个高于 3K 的温度而我要的是绝对零度，我们失去了机会……我的过错还在于后来又失去了一次，那是在 1961 年关于相对论的第 20 次瓦伦纳暑期讲习班上，当我和迪克作完全同样的讨论的时候。就微波背景而言，我显然是太拙于"发现"了……

那个 1961 年在瓦伦纳同霍伊尔讨论过的迪克，更是一位误失良机者的典型。除了那次讨论之外，他误失在历史上成为"测定宇宙温度的人"的机会不是一次而是两次，第二次时他竟忘记了自己以前关于这个问题的工作！迪克于 1916 年出生在美国密苏里州的圣路易斯，比霍伊尔小 1 岁，在 20 世纪 30 年代后期毕业于普林斯顿，1941 年在罗切斯特大学获得博士学位，第二次世界大战时在麻省理工学院研究雷达，1946 年到普林斯顿任教，此后就一直在那里，还当了物理

系主任和爱因斯坦科学教授。迪克是别人骗不了的，但他在 20 世纪 40 年代也没能看出现在事后看来似乎很明显的东西。

迪克在麻省理工学院曾制造过一种仪器，用于测量电磁谱微波部分波长很短的射电辐射。这种仪器现在就叫迪克射电仪，其原理仍被用于派同样用场的现代仪器中。他和三个同事把这种仪器中的一台朝向天空，想看看是否有来自河外星系的微波辐射背景。表示这种辐射强度的一种方式是用温度。迪克他们得出，有一种温度在 20K 以下的背景辐射，这是他们的仪器所能给出的限度。他们写了一篇文章报告这个结果，并在《物理学评论》上发表，在那家刊物的同一卷里还登有伽莫夫 1946 年的那篇关于核合成的论文。迪克小组的文章在前(第 70 卷 340 页)，伽莫夫的稍后(第 70 卷 572 页)。两篇文章的写作动机毫无关联，却在同一卷刊物上出现。20 世纪 50 年代的每一个学生或更年长的研究者，查找伽莫夫的这篇文章大概会是循踪自 αβγ 论文以来或自阿尔弗和赫尔曼的工作以来的描述，从而觉得手中拿的确实是宇宙火球曾经存在的证据。而霍伊尔、伽莫夫或是某个无名学生，在查找伽莫夫的这篇文章时，应该会碰上迪克小组的文章并因而作出点推理，但这种事从未发生。有时科学发现似乎有自己的意愿，要等到时机成熟时才出现。

到 20 世纪 60 年代初时，迪克自己已经完全忘记了那次测量，但他的思想却转向了宇宙学。他令人吃惊地似乎完全不理伽莫夫、阿尔弗和赫尔曼的开创性成果，自己独立研究这样一个模型宇宙，它从很大的尺度坍缩成为火球，然后从高密态反弹并膨胀。迪克所迷恋的主意是，现在的宇宙是在永久振荡的膨胀阶段，宇宙的每次循环都是膨胀后收缩，收缩后反弹并重新膨胀。他需要坍缩来产生高温高密的状态，然后才有反弹；在坍缩的宇宙里所有物质都终将分解成中子和质

子，然后是新一轮膨胀开始。没有任何“信息”能从宇宙的一次循环传到下一次，于是对任何一个生活在膨胀宇宙中的人来说，宇宙就正像是从大爆炸中产生出来一样。

这一切在很大程度上仍是把宇宙学作为一种游戏、一种智力练习。但是迪克作为一个观测者的经验(尽管是已被部分遗忘的经验)使他和他的同事最后还是走上了正确的道路。他交给普林斯顿的一名年轻研究者皮布尔斯(P. J. Peebles)一项任务，算出在那个振荡模型里宇宙的温度如何随其演化而改变。皮布尔斯自己也不知道他是在重复阿尔弗和赫尔曼 15 年前就已做过的计算，他得出，如果我们生活的宇宙是从热大爆炸开始的，它现在就应该充满着一种温度约为 10K 的背景辐射。迪克于是在 1964 年鼓励普林斯顿的另外两个研究人员去寻找这种辐射。罗尔(P. G. Roll)和威尔金森(D. T. Wilkinson)就装置了一台探测器(迪克射电仪的一种变体)，并开始在普林斯顿物理实验室屋顶上装一个小天线，以探测那温度只有几开的宇宙背景辐射。就在已经迈出划时代发现的最后一步时，普林斯顿小组脚下的地毯被抽掉了。迪克接到了贝尔研究实验室一个年轻人的电话，那实验室是在新泽西州的霍尔姆德尔，离普林斯顿仅 30 英里(约 48 千米)。打电话的人叫彭齐亚斯(Arno Penzias)，他和同事威尔逊(Robert Wilson)用自己的射电望远镜，也就是一个以前在通信卫星实验中用过的 20 英尺(约 6.1 米)号角状天线，得到了一些很古怪的结果。有人提议迪克可能能解释这种费解的宇宙背景辐射，大家还是一起来讨论一下……

创世的回声

彭齐亚斯来自慕尼黑的一个犹太人家庭。他生于 1933 年，正好是盖世太保成立的那一天(4 月 26 日)。他家是 1939 年离开纳粹德国

去英国的最后一批家庭之一，他和弟弟是春天去的，然后是父亲，最后是母亲。全家团聚后于1939年12月乘船去纽约，1940年1月到达并定居下来。教育使这个已经穷困的移民家庭的儿子有机会在世界上发展，彭齐亚斯于1954年从纽约城市学院毕业，得了一个物理学学位。作为通信兵在军队服役两年后，他进了哥伦比亚大学当研究生，1962年获博士学位。

彭齐亚斯在哥伦比亚大学的导师汤斯(Charles Townes)是一位对微波激射和激光的发展有关键性贡献的物理学家，1964年获诺贝尔奖。[13]微波激射能用作探测弱射电发射的放大系统的基础，彭齐亚斯造了一个21厘米波长上的微波激射接收器，希望能探测到星系际氢，因为氢气体的辐射就在这个波长上。但他失败了，主要是因为并没有什么星系际氢可观测，他只能给出一个星系之间可能有的氢数量的上限。这种“失败”对一个博士生，甚至对其他大的研究规划来说都不罕见，哥伦比亚的考官们显然很高兴彭齐亚斯能有效地工作，尽管得到的是否定的结果。但是他对自己工作的看法就要严厉得多。他对《零上三度》一书的作者伯恩斯坦(Jeremy Bernstein)说过：“我在哥伦比亚侥幸过关……那是一篇糟透了的论文。”无论是否糟透了，同汤斯一起工作和在射电天文学上的首次尝试都深刻地影响了彭齐亚斯此后的生涯。

汤斯是1948年从贝尔实验室来到哥伦比亚大学的。贝尔实验室原是贝尔电话公司的一个研究室，后来被AT&T吞并，最近的反垄断法规要将联合大企业分解，这使得该实验室能否继续作为一个独立的研究机构有了不确定性。无论今后命运如何，贝尔实验室有足以自豪的研究历史，其中就有射电天文学的奠基人央斯基(Karl Jansky)在20世纪30年代对来自天空的射电噪声的首次发现。由于汤斯与贝尔实

验室保持着联系，彭齐亚斯于1961年即获博士学位前不久进入了坐落在霍尔姆德尔附近克劳福德山上的射电研究实验室。贝尔实验室虽然基本上是为母公司的利益搞实用研究，但也总保持学术研究的传统，这就使得第一流的科学家都愿意加入，也使得实验室的实用研究与大学和其他学术机构取得的新进展之间能充分沟通。彭齐亚斯起先是研究有关卫星通信联系的问题，后来获准回到射电天文学，但在另一个未来的射电天文学家威尔逊参加进来之前进展甚微。实验室里只有一个射电天文学职位，于是就由他们两人分担，每人都用一半时间搞射电天文学，另一半时间去干别的。

威尔逊来自一个与彭齐亚斯大不相同的背景。他1936年生于得克萨斯州的休斯敦，父母都上过大学，父亲是化工工程师。他在休斯敦的赖斯大学上学时所有科学课程的成绩都是A，1957年毕业时美国最好的两所研究型大学，麻省理工学院和加州理工学院，都愿意录取他当研究生。他选择了加州理工学院，但对作什么方面的研究却没有拿定主意。这时他受到两位英国天文学家的影响。一个是霍伊尔，在加州理工学院以访问教授的身份讲授宇宙学课程，使威尔逊喜欢上了稳恒态理论；另一个是德怀斯特(David Dewhirst)，建议威尔逊去同一位当时在该学院的澳大利亚射电天文学家博尔顿(John Bolton)一起作研究。于是威尔逊就去和博尔顿一起对银河系作射电巡视，绘出其中的氢气体云。结果倒并不是特别重要，氢气体云图是绘出来了，只是证实了澳大利亚一个小组绘制的一幅同类图的准确性。威尔逊也像彭齐亚斯一样对自己的第一个研究课题的完成质量表示不满："坦率地说，我认为我的论文没有多大科学产出，尽管它是一个很好的练习，而且我有机会见到了世界上大多数射电天文学家，他们来学院访问过。"[14]彭齐亚斯在完成博士论文前不久就离开了哥伦比亚大学，

而威尔逊则在1962年拿了博士学位后还在加州理工学院待了一年。于是在1963年，当威尔逊得知贝尔实验室对射电天文学的关注，而且克劳福德山上仍然比较新的号角状天线可以使用时，就断然决定去新泽西和彭齐亚斯合作。

那架天线曾用于“回波”系列卫星的工作。那些卫星只是送到轨道上的一些大金属球，用于反射世界各地的射电信号。它们只是像天空中的镜子，本身没有放大作用，所以当信号返回到地面站时就相当弱，就需要一个很好的天线系统来捕捉，并且要充分放大后才能有用。后来有了通信卫星，能把接收到的地面信号放大后再发回到别的地面站，克劳福德山天线原定的作用也就完结了。彭齐亚斯和威尔逊才获准把通信接收器拿掉，把天线变成一架射电望远镜，这需要几个月时间。他们想要把新接收器做得尽可能地灵敏，能探测非常弱的天文射电噪声。这样他们就得先尽自己所能消除所有用于放大空中射电波的电器系统本身产生的噪声。这有点像调幅收音机里的静电干扰，那些嘶嘶的背景噪声有的来自偶遇的射电波（包括来自天空的射电波），有的则是由于接收器本身的效能不够。这种静电干扰或背景噪声可以用温度这个量来衡量，而那些把克劳福德山天线用于“回波”卫星工作的工程师已注意到，整个系统的静电干扰比他们能找出原因的还要稍多一点，或者说，天线温度过高。一位名叫欧姆（E. A. Ohm）的工程师，曾在1961年的《贝尔系统技术杂志》的一篇文章里报告过，在减去了所有能够解释的噪声之后，还有相当于温度约为3K的额外辐射噪声。这并不足以破坏“回波”系统，所以工程师们也不怎么担忧。但彭齐亚斯和威尔逊却必须在开始既定的射电天文研究之前跟踪和消除它，至少也要能加以识别。

于是彭齐亚斯和威尔逊就忙于探查自己系统中那令人恼火的噪声

源。他们做得如此之彻底，甚至去清扫那号角状天线上的鸽子粪，但都没有效果。与此同时，普林斯顿小组正在平静地制造仪器以探测宇宙背景辐射。也是在那个1964年，在远处的英国，霍伊尔（又是那个人！）和泰勒（Roger Tayler）正开始沿同一条路线前进，即着手计算大爆炸宇宙今天的背景温度。而在更远的苏联也起了一阵疾风。泽尔多维奇（Ya. B. Zel'dovich）的计算表明，要解释宇宙中观测到的氢、氦和氘的丰度，宇宙就必须从热大爆炸开始，并且现在有一个几开的温度；他甚至知道《贝尔系统技术杂志》上欧姆的那篇文章，但却误解了其中的专门术语，以为欧姆的测量表示宇宙的背景温度在1K以下。苏联的另一位研究者斯米尔诺夫（Yu. N. Smirnov）计算出背景或残余辐射的温度是在1—30K范围内。依据这个计算结果，多罗什克维奇（A. G. Doroshkevich）和诺维科夫（I. D. Novikov）还写了一篇文章，从微波背景的角度来讨论已有的各种射电天文测量的作用。他们得出结论，对探测这种背景来说当时世界上最好的天线就是克劳福德山上的贝尔实验室天线，并建议把那架天线用于此项目的。所有这些工作都是在1964年完成的，并且大多数在当年发表。时机已经成熟，宇宙微波背景该向世人露出真容了。尽管至少有两大洲上这4个研究组在朝同一目标逼近，彭齐亚斯和威尔逊却仍然解不开仪器系统的额外噪声从何而来这一难题。

关于难题最后如何解开的各种叙述稍有不同，但基本上是一致的。按照其中一个版本，彭齐亚斯1964年12月去蒙特利尔参加一个天文学会议，在返回的飞机上坐在麻省理工学院的伯克（Bernard Burke）旁边。在飞行中彭齐亚斯谈到了正在和威尔逊一起消除他们系统中的背景噪声。几天后伯克打电话给彭齐亚斯，建议他去找普林斯顿小组。故事的另一版本则是，彭齐亚斯在为别的事打给伯克的电话

里偶然提到了背景噪声。不管是哪种说法，总归是在 1965 年 1 月的一次电话里，麻省理工学院的伯克告诉克劳福德山上的彭齐亚斯，另一位天文学家即华盛顿特区卡内基学院的特纳(Ken Turner)曾听过普林斯顿的理论家皮布尔斯的一个报告，其中预言有一种电磁辐射的背景噪声充满宇宙，辐射温度大约是 10K。按照伯克的建议，彭齐亚斯给迪克打了电话，普林斯顿小组的全部 4 名成员很快就驱车半小时来到了克劳福德山，要看看究竟是怎么回事。理论和观测终于会合了，2 加 2 的确等于 4。

普林斯顿小组对这项发现要比彭齐亚斯和威尔逊激动得多。对普林斯顿研究者来说观测恰与理论(他们认为是自己的理论)预言相一致，这是科学方法成功的一个很好例证。对彭齐亚斯和威尔逊来说，他们所测量的射电噪声有了一个解释固然可以宽慰，但似乎仍可能有别的解释。此外，在更多的证据到来之前威尔逊也不愿接受稳恒态假说的失败。尤其是，测量仅在一个波长即 7 厘米上进行，必须用不同的接收器对许多其他波长重做，然后才能真正认识背景辐射的本质。

所以这消息虽然很快就在学术界传开，却是以一种十分慎重的形式印成了文字。两个小组商定各写一篇文章提交《天体物理学杂志》一起发表。相比之下，普林斯顿的那一篇要更有趣和令人兴奋得多，也登在前面(第 142 卷 414 页)；彭齐亚斯和威尔逊的文章紧随其后(第 142 卷 419 页)，用了个很乏味的标题“4080 兆赫上天线额外温度的测量”，文中只有一句话涉及这项使他们两人在 1978 年获诺贝尔奖的发现的真正意义：“对所观测到的额外噪声温度的一种可能解释，由迪克、皮布尔斯、罗尔和威尔金森在本卷中与本文配对的一篇快讯中给出。”但那卷《天体物理学杂志》最惊人之处也许是那两篇文章都没有提到伽莫夫、阿尔弗和赫尔曼的工作。这个失误很快就被改正了，

后来的有关出版物全都给了那些先驱者应有的荣誉，但在此之前他们的工作曾长期遭受冷遇。

后来在各种不同波长上的测量确凿无疑地表明，彭齐亚斯和威尔逊所称的“额外噪声”的确是宇宙背景电磁辐射，而且正好就是关于我们宇宙起源的大爆炸模型所要求的温度近乎 2.7K 的“黑体”辐射。[15]这真正就是创世的回声，是我们用仪器能够得着、摸得到的大爆炸残余。这个发现应当列于有史以来最重要的科学发现之中，它改变了宇宙学的面貌，使宇宙学家认识到，他们不是在玩什么智力游戏，他们手中的方程式确实能够描述我们宇宙的起源和其中的一切。随着对这种剩余辐射的认识，“我们是从哪里来的？”这个问题从哲学的领域转到了科学的领域。这就是为什么伽莫夫和他的同事们超越了他们的时代——因为在 20 世纪 40 年代和 50 年代时只有他们相信那些方程。当宇宙学的确是一门科学这个认识随着背景辐射的消息传播开时，有不少物理学家转到了宇宙学，其中一位是温伯格，他作了恰当的概括：

> 伽莫夫、阿尔弗和赫尔曼应该首先享有殊荣，因为他们愿意认真地看待早期宇宙，因为他们用已知的物理定律揭开了最初三分钟的奥秘。是的，他们没能走出最后一步，就是使射电天文学家认识到应该去寻找一种微波辐射背景。1965 年 3K 辐射背景的最终发现带来的最重要结果，就是使我们全都认真地接受了曾经有过一个早期宇宙的思想。[16]

勒梅特在 1966 年去世前不久才听到消息。伽莫夫也只比他多活了两年。假如他们活得稍长些，或者背景辐射发现得稍早些，他们应该能分享诺贝尔奖，因为那个发现所证实的大爆炸概念是他们提出

的。诺贝尔奖从不授予死者，当诺贝尔奖委员会于 1978 年认定将早期宇宙的真实性记录在案的时候已经来到时，他们面临着一个看来很棘手的问题——授给谁？候选者并不缺乏。一方面是一对年轻的射电天文学家，他们发现了古怪的现象却不知道是什么，甚至别人来告诉了那是什么之后，他们开始时还不大相信。另一方面是一个小组，他们预言了那个背景的存在，制造了仪器去探测，而且在克劳福德山那次至关紧要的会见后很快就用自己的仪器证实了自己的预言。除开所有那些虽靠近而未能命中者，如苏联人、霍伊尔和泰勒等等，还曾经有并仍然有第三只“手”应该考虑，就是阿尔弗和赫尔曼，他们是伽莫夫小组还活着的成员，他们首先说出了这一切，尽管曾被忽视。

获奖的是彭齐亚斯和威尔逊。在那种情况下，也很难再授给别人而不显得想要的太多反而成为笑柄。或者有办法做到？我在想委员会是否考虑过，哪怕是用片刻时间考虑这样一个有灵感的决定：为什么不能把奖授给第一个报告探测到 3K 背景的欧姆呢？他不知道自己发现的是什么，但是彭齐亚斯和威尔逊也不知道，而毕竟欧姆的确发现在先。

这些议论都没用了。做过的已经做过了，不可能重做了。连宇宙看来都是如此。它从大爆炸开始，此后就平稳地演化至今。有了今日宇宙的温度这样一个量来校准大爆炸，宇宙学家就能使他们的计算更精确，从而得出创生的标准模型，即宇宙从创生后不到 1 秒时起直到今日的整个故事。然后他们就能进而攻坚，探索到那创生本身的第 1 秒钟之内，探索大爆炸之前的宇宙那个奇怪的世界。

注　释：

1. 可以用弹簧做类比而使这一点更形象化。假如你想把太阳再分开成稀云，你就必须做功来克服引力而使每个粒子都离质量中心更远。稀疏状态的粒子就把你做的功以引

力势能的形式储存起来，而这正是这块云收缩时以热的形式释放的能量。

2. 开尔文一亥姆霍兹时标的一个含义是，太阳对其内部产热过程的任何猛烈变化要用大约1000万年时间来作出调整。如果这种产热过程，且不管它具体是什么，突然停止了，地球上将要过好几百万年才能注意到这种变化，因为渐进的引力收缩会使太阳保持高温。这或许能使人对那些反常的时期感到一点安慰。

3. 在太阳这样的恒星内部，原材料全是氢核即质子。其半数会转变成中子，每个质子转变时都放出一个正电子以带走多余的电荷。正电子是电子的带正电荷的对应物，正电子与电子相遇时会湮灭，它们的全部质量都转变成能量。这就使得定量计算稍有不同也略微复杂一点，但结果仍是，每次制造一个氦核都有约0.7%的初始质量转变成能量。

4. 与他同时即1938年，另一位德国物理学家冯·魏茨泽克(Carl von Weizsäcker)也提出了同样的保持太阳高温的基本机制。像贝特一样，冯·魏茨泽克也在20世纪40年代研究了制造原子弹的问题，但他是在德国工作，这就可以解释为什么他没有在1967年与贝特分享诺贝尔奖，尽管据说他曾有意不让自己小组的工作为纳粹提供核武器。

5. 更准确地说，在1938—1939年提出的只是循环的CN部分；氧的作用较小，改名为CNO循环是后来的事。实际上这是两个循环，分别是CN和NO，一些有修辞癖的天文学家就将其另名为CNO"双循环"。

6. 第55卷第434页。

7. 重印于*Science*,Vol.226,p.922。

8. *Monthly Notices of the Royal Astronomical Society*, Vol.106,p.343.

9. 第29卷第547页。加拿大出生的美国天体物理学家卡梅伦(Alastair Cameron)独立地得出了类似的结果，并于1957年发表于*Publications of the Astronomical Society of the Pacific*, Vol.69,p.201。如果说这样在注释中提及对卡梅伦有失公正，要记住毕竟是霍伊尔给出了福勒在其诺贝尔奖演讲中所称的"恒星中元素合成的主要概念"，而且是福勒小组作了反应速率的实验室研究。历史给予B^2FH以高位也许是对的，尽管有人会认为伯比奇夫妇很幸运，在正确的时候和正确的地方参加了福勒和霍伊尔的工作。

10. 福勒和霍伊尔在学术上几乎是不可分离的。福勒迄今已发表了200多篇论文，其中不少于25篇即约10%也有霍伊尔的名字。诺贝尔奖委员会没有看到应该在1983年把这两个名字保持在一起才合适，这的确是件令人遗憾的事。

11. *Physics Bulletin*, Vol.35,p.17.麦克雷肯定知道他在说什么。他同霍伊尔一样也是稳恒态模型的创立者之一，而且作为一名深刻地思考科学研究本质的优秀学者，他仍在告诫大爆炸模型并没有被"证明"正确。简单的大爆炸模型和简单的稳恒态模型代表着一个巨大的可能范围的黑白两个极端，两者之间还有各种各样的灰颜色。他说，重要的是以开放的思想来讨论所有的可能性，即使一个模型看来是对今日世界的比其他模型更好的描述，任何一个科学家都仍然不大可能以绝对的信心说"这是对的"。但是热大爆炸模型现在已被绝大多数天文学家接受为最好的宇宙模型。霍伊尔是稳恒态模型的主要提出者，却又是那篇表明大爆炸模型正确性的里程碑式论文的作者之一，这倒真是一种讽刺性的幽默。

12. G. Herzberg, *Spectra of Diatomic Molecules*, 2nd ed. (Princeton, N. J.: Van Nostrand,1950).

13. 微波激射是"微波辐射的受激发射放大"这个词组的缩写；对本书来说重要的词是两个，一是微波，即波长为几个厘米，二是放大，即能使弱的射电输入加强。激光也是

同样一回事，只不过是可见光而不是微波。

14. 亦见伯恩斯坦的书。

15. 但仍然可能对背景辐射作出另一种解释。哈佛大学的莱泽(David Layzer)最近大事声张地争辩说，背景辐射可能是由在星系之前形成的一代大质量恒星产生的，而宇宙在开始时是冷的。这样的可能性肯定值得讨论，并且可以有这样的提醒作用，即公认为“最佳选择”的宇宙模型也不一定就代表最后结论。莱泽在《构造宇宙》(*Constructing the Universe*)一书里提出了他的宇宙学版本；遗憾的是(如果可以理解的话)他没有持公平态度，而是把自己的版本作为“最佳选择”，过急地把标准模型扔在一边。

16. *The First Three Minutes*, Deutsch ed., p. 132.

第六章　标准模型

我们生活在一个膨胀着的宇宙里，它均匀地充满着微弱的电磁辐射，物质以团块的形式(从大尺度上看)均匀地散布于各处。为着了解宇宙在很久以前的状况，我们就得想象着把钟拨回，因而把模型宇宙缩小。模型宇宙缩小的结果是其密度增大。物质密度的增大是因为同一总量的物质随着时间反演而被挤进较小的体积里，辐射的密度也同样增大。辐射密度的增大表现为蓝移，即辐射的波长缩短，也可以借助于温度来表述，即从今天的3K开始，随着时间的回溯而变得越来越热。

从现在起我将讲述宇宙的标准模型，即目前宇宙学的“最佳选择”，把它作为真实宇宙的写照。这是一个必要的声明，因为从现在起那种故事性的叙述将被打断，将不再总是提醒说这只是我们已有的对宇宙的最好描述，而新的进展可能取代它。如果一个模型被严格地提出来并且用方程式表达出来了，我们就能说在这个模型里随着其演化必定会发生什么。在标准模型里必定发生的事情的主要特征，看来

与我们在宇宙中能看到的很相像，所以标准模型是一个好模型。我们也希望，在我们不能直接观测的地方和时候，那些按照标准模型必定会发生而我们不能直接看到的事情也能告诉我们真实宇宙是什么样的。但是模型绝不能告诉我们在真实宇宙中必定发生什么。我将给出一个描述，把它作为我们宇宙的历史，这是唯一有条理的方式。但请保持这样的认识，即这个描述实际上只是对一个数学模型宇宙而言，这个模型宇宙与我们今天的宇宙有显著的相似性，所以我们认为能够用它来认识很久以前当我们的宇宙年轻时发生过什么。

时间上的回顾对应于模型宇宙的收缩，即目前膨胀的时间反演。如果模型宇宙的这种收缩持续得足够长久，那么按照物理定律就将成为一个物质和辐射密度都无穷大的点。但是，我们由地球上的实验和由对今日宇宙的观测得知的物理定律又显然不适于描述无限大的物质和能量密度，因而这些定律只能运用到返回创生时刻的想象旅程的某一阶段为止。如果我们暂且抛开那个在创生本身的第 1 秒钟之内精确地发生了什么的难题，那么对膨胀宇宙的观测又完全足以告诉我们，创世必定发生于 100 亿到 200 亿年之前。为便于论述，不妨取估计年龄范围的中间值，认为 150 亿年前就是 $t=0$ 的时刻，此时宇宙以一种极高密度和极高温度的状态出现，然后描述宇宙如何自 $t=0$ 起随时间流逝而演化至今。我们这样做是通过在想象中把时钟倒拨，从现在起拨回到尽可能靠近那无穷大能量密度的状态，然后再想象让时钟与宇宙演化进程一样走回来。所以，作为开始，我们需要检查一下对今日宇宙知道些什么。[1]

第一，我们知道它在膨胀。第二，我们知道恒星物质的大约 25% 是氦，其余的基本上是氢。第三，我们知道它由温度为 3K 的辐射所充满。电磁辐射也可以用粒子的概念来表述，那就是光子。宇宙中每

立方厘米的空间大约有 1000 个光子。如果按观测所启示的那样，假定常见物质的密度大致对应于一个平直宇宙，即在开放和闭合状态之间的分界值附近，那么宇宙中光子的数量大约是质子和中子总数的 10 亿倍。这种辐射的能量大体上是由 $E=mc^2$ 给出的宇宙可见物质能量的 1/4000。今日宇宙是由物质主宰，但并非历来如此。

倒拨时钟，我们就能由简单的物理定律计算出辐射温度如何随宇宙的缩小而升高。每个质子或中子所具有的能量是不变的，但每个光子所具有的能量就随着辐射被挤压和蓝移而增大。当温度约为 4000K 时，每个光子的能量是每个质子或中子能量的 10 亿分之一，但由于光子的数量多上 10 亿倍，辐射的总能量就与物质的总能量相当了。对所有更高的温度，相应地也就是更早的时期和更大密度的状态，宇宙是由辐射主宰的，物质只起次要作用。于是我们就有了重新构造出大爆炸状态的公式，因为我们知道今天宇宙的温度，知道每个质子或中子所对应的光子数，还有关于状态如何随时间倒流而改变的定律。那些定律的最简单形式即弗里德曼—勒梅特宇宙学，加上关于今日背景辐射的已知事实，就是在宇宙演化的最初几分钟里把 25%的原始物质烹制成氦的秘方。

宇宙火球

在对宇宙作有效的描述时，我们能把物理定律推回去多远，也就是离 $t=0$ 多近呢？今日世界上自然存在的物质最大密度是在原子核里，在那里质子和中子紧紧挤在一起。核反应(即涉及质子和中子的反应)是我们在身旁看到的各种化学元素之所由来，而在宇宙诞生后很快就发生的核反应确定了后来成为第一代恒星燃料的氢和氦的比例。标准大爆炸模型得益于瓦戈纳、福勒和霍伊尔对大爆炸能产生多

少氦所作的计算，能够有效地列出宇宙自其物质密度与今日原子核物质密度大致一样或稍低一点的状态以来的演化史。那种状态所对应的温度约是 10^{12}K，密度约是核密度即 10^{14}克/厘米^3，时间则是在 $t=0$ 之后的 10^{-4}秒。

这些条件是如此之极端，在细述宇宙如何从这种状态即从大爆炸本身演化之前，有必要提一提描述这种极端条件的有关物理定律。至关重要的一点是，辐射在大爆炸中起着比在今日宇宙中重要得多的作用，这个道理很容易明白。如果让模型宇宙随时间倒退而收缩，那么在很长时间里对单个原子来说并没有什么重要的事情发生，更不用说原子核了。这是因为星系之间相隔如此之远，要经过长达数十亿年的收缩才会彼此接触。即使到了那时候，还要经过很长时间的收缩才能把单个恒星挤压成一种无定形的团块。但背景辐射，尽管只是一种温度仅为 3K 的微弱嘶嘶声，却充满着今日的整个宇宙，并且总是充满整个宇宙。想象中的收缩一开始，辐射就会增强，温度就会升高。到恒星被挤压到单个原子核开始受影响时，空间每一点的辐射密度就已经增大到其能量远大于粒子所具有的能量。这时的辐射就不再是“背景”，而是一个又热又密的宇宙中进行的物理过程的最前线。

一个质量为 m 的粒子所等价的能量当然是 mc^2，而能量为 E 的辐射所等价的质量也简单地就是 E/c^2。$E=mc^2$ 这个关系式表明，一个足够高能量的辐射包(即光子)能转变成具有相应质量的物质，反之亦然(还有其他规则即量子规则必须遵守，下文很快将会谈到)。在大爆炸的高能量高密度情况，用粒子的概念来考虑辐射的确是有意义的；如我们在下面很快就会看到的，事实上在量子物理的奇特世界里所有的粒子也都能被看成波，而所有的波也能被看成粒子。能量和质量是等价的和可以互变的，粒子和波的概念也是如此。但是一个光子即一

个能量包，不可能简单地为一个粒子所取代而消失。粒子都是配对的，每种粒子都有其对应物，叫做反粒子，表示是粒子的“镜像”。电子的镜像是叫做正电子的粒子，带有正电荷而不是电子所带的负电荷，故得其名。如果一个电子与一个正电子相遇，两者就会湮灭而成为高能辐射，即 γ 辐射。而足够高能量的 γ 辐射也能变成一对粒子，即一个正电子和一个电子。

在我们现在所谈的时间，即大爆炸的 $t=0$ 之后 10^{-4} 秒到 0.1 秒之间，宇宙由辐射主宰。可以从两个途径来设想这种主宰。第一，辐射的密度（每个小体积里所包含的能量）高到如此程度，即（粗略地讲）在每个相当于电子—正电子对尺度的空间体积里就有等价于一对电子—正电子的能量。于是能量就能很容易地由电磁能转化成电子和正电子并且又可以转变回来，同样也可以用电磁能粒子即光子的概念来思考。对应于每一个核粒子（每个质子和每个中子）都有 10 亿个光子，那些光子中的每一个都能够、而且确实会变成一个正负电子对；而在这个原始大动荡中的电子和正电子遇上时，又会湮灭而生成新的 γ 光子以代替已变成正负电子对的光子。火球是由光子、电子、正电子，还有一种叫做中微子的无质量粒子所支配。或许“火球”不是对这时的宇宙的最好称呼。按照公式 $E=mc^2$，$t=0$ 后 0.01 秒时宇宙的能量密度相当于地球上水密度的将近 40 亿倍。好一个火球！

质子和中子（合称为核子，因为它们都是核中的粒子）即使在这样极端的条件下也相对较稳定。中子在单独存在时会在几分钟里自发衰变，变成一个质子、一个电子和一个中微子。但是火球的时间尺度只是 1 秒的很小部分，所以一个能稳定几分钟的粒子就等效于是永恒的。质子和中子的质量相近，即都稍小于电子质量的 2000 倍。要产生质子—反质子对或中子—反中子对，相应地就需要更高的辐射能量

密度(更高能的光子)。在宇宙的更早时刻即 $t=10^{-4}$ 秒之前可以得到所需的能量，但是20世纪60年代后期的标准大爆炸模型只涉及宇宙密度降到低于核物质密度后的事情，那时质子和中子已经从辐射中凝聚出来。

在讲到随着宇宙的进一步冷却将会发生什么之前需要强调的最后一点，是关于所有这些变化所发生的时间尺度。今天的宇宙作为一个整体不会在0.0001秒内，甚至不会在1000万年内有显著变化。宇宙学家说宇宙年龄是100亿年到200亿年之间，他们对这个取值范围有2倍的可能差异并无不安。但当宇宙年轻时状态就改变得更快，1秒的很小部分对大爆炸时的事件都很重要。在宇宙演化的任何阶段的特征时间尺度，都可以取成宇宙中任一选定区域的大小加倍所需的时间(今天这就等价于任何两个星系团之间的距离加倍所需要的时间)。引力在持续地使宇宙的膨胀减慢，所以这个时间尺度本身是随时间增大的。使宇宙的整体面貌发生可观变化的时间越来越长。反过来必然是，越接近 $t=0$，任何重要变化所需的时间就越短。宇宙在任一时期的特征时间尺度粗略地反比于其密度的平方根(密度越大，特征时间越短)，在光子、电子、正电子和中微子主宰的时代开始时，这个时间尺度是0.02秒。

要大略感受一下这个重要的时间尺度如何随宇宙年龄变化，可以只取10的幂并沿时间反推。宇宙年龄是大约150亿年，若取10的整数幂就是 10^{10} 年。天文学家对他们所作的宇宙年龄估计是感到高兴的，因为所有的估计都与10的同一个幂相符，没有一个估计小到 10^9 年或是大到 10^{11} 年。如果回顾宇宙的过去，第一个重要的里程碑是在大约 10^9 年，那时的宇宙是它现在年龄的1/10，看去应该是显著不同。下一个里程碑将是它再年轻10倍即 10^8 年时，也就是现在年龄

的1%，如此等等。这样看来，在从第1个1/10秒(0.1秒)到第1秒末的间隔里所发生的一切，与在从第1个1%秒(0.01秒)到第1个0.1秒末的间隔里所发生的一切大体上是同样有趣和重要，依此类推。这种类比并不精确，但却给出了早期宇宙迅速变化的一点味道。还有另外一种透视方法。宇宙年龄若以秒为单位是数倍于10^{17}。1秒就是10^{0}秒，从现在到第1秒的间隔覆盖的广度是10的17次幂。如果再往第1秒的另一侧作同等程度的回溯，就会到达10^{-17}秒。在很真实的意义上，从10^{-17}秒到1秒的间隔等价于从1秒到现在的间隔。物理学家现在谈论着发生在$t=0$后10^{-40}秒内的事情，也就是说从$t=1$秒向创世时刻的回推已是第1秒与现在间隔的2.5倍。如此说来，从10^{-4}秒到大约4分钟之间的事件几乎就在眼前，正是这段时间的事件使我们的宇宙定形。

最初四分钟

对宇宙火球从辐射、正负电子对和中微子支配的时期起往后作出最好描述的是温伯格的书《最初三分钟》。如温伯格在书中所声明的，这个标题稍有点作者自己的随意。他所讲的大爆炸实际上是从10^{-2}秒即创世时刻后的1%秒开始，他所描述的主要活动是发生在此后的3分46秒时间内。他写这本书是在1976年，当时的物理学对第1个1%秒内发生的事情还迷惑不解，所以他的起点是选得很合理的。这本书现在仍然是对那个极为重要的3¾分钟里的情况的清晰指南。[2] 所以我也就照搬温伯格对4分钟内的状态变化所作的经典性概览，在这段时间里宇宙从一团均匀而密集的辐射和物质的汤，转变成约75%的氢和25%的氦的混合，而辐射则与物质分离，并逐渐衰退而成为我们今天所知道的背景。

故事从时间 $t = 10^{-2}$ 秒、宇宙温度为 10^{11} K 开始。那时主宰宇宙的是辐射、由辐射所产生并能重新湮灭成为辐射的正负电子对、还有无质量的中微子及与之配对的反中微子。质子和中子对于今天的物质是如此重要，是它们组成了所有的恒星和行星、太空中的气体和尘埃云乃至我们自己体内的原子，但那时只是那团汤里无足轻重的成分，它们的数量只是光子数的 10 亿分之一。它们不断地遭受电子、正电子和中微子的轰击，因而不断地改变面目。一个反中微子与一个质子碰撞能生成一个正电子和一个中子，而一个中微子与一个中子碰撞又能生成一个电子和一个质子，这两种反应又都能反向进行。单个的核子被不断轰击，反复地由中子变为质子又变回来。但平均说来，只要火球的能量高得足以使所有这些反应很容易地进行，在宇宙的任何部分就总是有大致相同数量的质子和中子。但是，当温度降到约 3×10^{10} K 时，情况就开始变化了。

粒子物理学家常用电子伏特(即 eV)这个单位来量度能量和质量(若计及 c^2 因子这两者就是一回事)。1eV 是一个电子被加速通过 1 伏特的电势差所获得的能量。这是一个很小的单位。一个典型的可见光光子所携带的能量约是2.5eV，一个电子的质量是 510 000eV，即稍高于 0.5MeV。[3] 质子的质量是 935MeV，中子的质量几乎但不完全与之相同。这个“不完全”正是宇宙演化下一个阶段的关键。

当宇宙的温度高达 10^{11} K 时，每个电子、光子或其他粒子所携带的典型能量约是 10MeV，即 1000 万电子伏特。有的具有更高的能量，有的低些，但这是一个很好的平均值。这个值远小于核子的质量，这就是核子何以在那时能保持其为核子的缘由。而这个值又比正负电子对的质量大许多，这就是这种粒子对何以在那时能很容易地产生出来的原因。这个值还比质子与中子质量之差要大许多，因为这种

差额还不到1.3MeV。对一个具有10MeV能量的电子或中微子来说，是与一个质子还是与一个中子发生反应几乎没有差别，质子转变成中子和中子转变成质子这两个方向上的反应都同样容易进行。但随着宇宙温度的下降，每个粒子所携有的能量也成比例地降低。由于驱动反应的能量小了，质子与中子的质量之差开始显得重要，将较轻的质子变成较重的中子的反应就变得相对难以进行。如果一个能量足够高的电子与一个质子碰撞，这种“爬坡”反应仍能发生，但是能量足够高的电子变得越来越缺乏，其数量与将中子转变成质子所需要的能量稍低的粒子就相差得越来越大。

在$t=0$后的0.1秒，宇宙温度是3×10^{10}K，能量密度已降到水能量密度的3000万倍，膨胀速率也已大大减慢，宇宙的特征时间尺度是0.2秒，虽然核子与光子数量之比仍是1∶10亿，但中子数与质子数之比就不再是50∶50，而是38∶62。

约在$t=0$后的1/3秒，宇宙中发生了一个重大变化。在早期火球的高温条件下，粒子之间很容易发生许多反应，包括电子、正电子和中微子之间的互变，即正负电子对湮灭而生成中微子—反中微子对，以及反向转变，还有已经讲到的核子反应。但是中微子在任何我们视为正常的条件下都很难与其他物质作用。它们能直接穿过地球而不受什么影响，太阳核心的核反应产生的中微子流也径直穿过太阳而去，并不受什么影响。对中微子而言，通常物质都是透明的，这里的“通常物质”包括任何不如创世后1/3秒内那样极端的状态。从那个时刻或稍后起，中微子停止与电子、正电子或任何其他粒子反应，而是保持作为一种背景海(颇像宇宙微波背景辐射，但难以探测得多)充斥于宇宙之中，对宇宙的演化只起次要作用。[4]

于是在$t=1.1$秒时，温度降到了10^{10}K，密度降到了水密度的

380 000 倍，中微子停止与其他物质反应(或者说它们退耦了)，宇宙的特征膨胀时间也延伸到了 2 秒，而中子数与质子数之比则进一步变成了 24 : 76。随着温度继续降到 10^{10}K 以下，具有足以生成正负电子对能量的光子变得更稀少，因而在宇宙演化的这个阶段，正负电子对湮灭的速率就超过了新对的生成速率。

从现在起，原先那种听得使人喘不过气的演化进度已减慢到近乎熟悉了，已经能用整秒而不再是秒的分数，而且粒子和它们的反应已经与为今天的太阳和其他恒星提供能量的粒子和反应很相似。

到温度降为 3×10^{9}K 即 $t=0$ 后 13.8 秒时，再也没有新的正负电子对生成，原有的对则仍在湮灭。氘核(1 个质子加 1 个中子)能短暂地生成，但几乎是一生成就被其他粒子碰撞而又拆开。虽然还有中子在转变成质子，但由于能量降低这个反应也急剧减慢下来，核子中仍有 17%是以中子的形式存在。从创世起的 3 分 2 秒时，宇宙已冷却到 10^{9}K，终于可以与今日宇宙作某些比较了。太阳核心的温度是 1.5×10^{7}K，$t=0$ 后 3 分钟时的宇宙温度已只是 70 倍于此。几分钟前那么重要的那些粒子反应实际上已经停止了，现在宇宙已经老到使中子的自然衰变变得重要，从现在起的每 100 秒内，剩余自由中子的 10%将衰变成质子，而现在中子在核子中占的比例已降到约 14%。但它们不会灭绝，因为随着温度降得更低，氘核能够形成而且不易破裂了。

现在，伽莫夫及其同事们概略地，而瓦戈纳、福勒和霍伊尔详细地描述过的反应终于可以发生了。核合成很快造出了氦 4 的核，但也基本上就到此为止，因为如前所述，不存在质量为 5 或 8 的稳定核，而如霍伊尔在 20 世纪 50 年代初所发现的，只有在恒星内部的条件下核合成才能越过那些缺口，而恒星现在还远未形成。[5]

一旦氦开始生成，所有可用的中子很快都被这样束缚起来而成为

稳定的了。这发生在中子在核子总数中所占比例在约13%或14%时，而在氦4核中每个中子有一个质子伴随。于是核子总质量中转变成氦4的比例就简单地是反应开始时中子丰度的2倍，也就是26%到28%之间。核合成是在创世后3分46秒开始的，当时温度为9×10^8K。到$t=4$分时，标准大爆炸模型已经造就了正好产生今日宇宙中观测到的那么多氦的条件。

标准模型的这个巨大成功所依赖的关键是，质子转变成中子和中子转变成质子的反应是在进行中"冻结"的，所以那剩余的14%或15%的中子能在核反应开始时留下来。那些使质子和中子互变的重要反应和反应的冻结点，都不仅对温度而且对早期宇宙温度下降的速率非常敏感。如果冻结是发生在"宇宙年龄"还只有几秒钟时，那么宇宙中氦所占的比例就会接近30%。但若一切都发生得稍稍再快一点，而冻结是在0.1秒时，由大爆炸产生的氦的比例就会几乎是100%(因为核合成也进行得快得多)，反之若宇宙的演化慢得多，因而冻结是发生在100秒时，大爆炸中就不会有氦生成，因为在核合成开始之前所有的中子已经转变成了质子。

温度下降的速率由标准模型以一种最简单的形式给定，并且与今日宇宙微波背景的温度相配合，从这种背景温度得出的一个至关重要的估计是，宇宙中每一个核子都对应着10^9个光子。这个比数在宇宙膨胀过程中保持不变，在光子已散布在宇宙中而核子却聚集成物质团块的今天是如此，在宇宙的火球阶段必定也是如此，那时的辐射主宰着物质，并且驱动着正好生成如我们看到的那么多氦的反应。标准模型还制约了今日宇宙中其他粒子存在的可能性。中微子和反中微子也参与质子转变成中子和中子转变成质子的过程，所以下文就要讲到，标准模型的成功还在于能给出大爆炸时业已存在、因而留待我们现在

去发现的中微子的数量和种类。

所以，继 20 世纪 20 年代对宇宙膨胀的惊人发现之后，标准模型已经有了第二个成功的预言，或是要求。宇宙中的氢是氦的 3 倍，这是大爆炸这个最简单的宇宙模型的一个特征性要求。标准模型三脚架的第三只脚是宇宙微波背景辐射。但要看清楚这种辐射的由来，从创世时刻开始的行程就不能再是以几秒或几分钟迈步，而是要一步跨越数千年甚至 10 亿年了。

后来的 100 亿年

$t=0$ 之后的半小时稍多一点(准确地讲是 $t=34$ 分 40 秒)，几乎所有的电子和正电子已经湮灭，这时的宇宙很像我们今天所知的真空。几乎所有的物质，但不是全部，已经消失。除了数量为光子的 10 亿分之一的核子外，在正负电子对最终湮灭时每 10 亿个电子中还有一个幸存下来，这正好是平衡宇宙中质子的正电荷所需要的数量，从而保证了物质最终成为电中性原子这种稳定的形式，即每个原子核中的每个质子都有核外一个电子的云相配。这个微小部分的物质是从哪里来的呢？为什么粒子与反粒子之间并不是完全对称并使一切物质都湮灭掉而只有辐射留存于冷却的宇宙中呢？答案来自对这样的粒子物理世界的认识，这个世界的状态甚至比宇宙在第一个百分之一秒后更加极端，这也是将在第三篇中描述的发现所解开的最容易提出却最有深远意义的难题之一。现在且只看创世半小时后的这个膨胀着和冷却着的火球。

现在火球温度已降到 3×10^{8} K，宇宙能量密度已只有水质量密度的 1/10。约 69%的能量由光子携带，而 31%是属于中微子，此时相应的膨胀时间尺度已伸长到 75 分钟。虽然所有可得到的中子已被装

进氦核里，宇宙仍然太热而不允许稳定原子形成，因为一旦一个带正电的质子或氦4核粘上一个带负电的电子，那电子又会被高能光子撞掉。这是宇宙的“辐射时代”，残余的物质由辐射支配着，没有什么重要的粒子反应需要操心。这个时代约持续700 000年，直到温度降至约4000K时为止，那时的核和电子就终于能经受住光子逐渐减弱的打击而保持在一起了。

原子形成的时间不能精确确定。早至创世后300 000年时就有一些氢原子开始形成，并能存活一段时间而不被辐射所电离；在 $t=10^6$ 年后，几乎所有电子都已被束缚在原子里，实际上相对于每100 000个稳定原子只有1个电子和1个质子还自由存在，物质完成了从辐射的“退耦”。从这时起辐射与物质之间就几乎不再有什么相互作用，因为电磁辐射与带电粒子之间虽有很强的作用，像原子这样的中性粒子与辐射之间就几乎没有什么相互影响。就像早些时退耦的中微子海一样，光子也离开并减弱成为一种宇宙背景。

这个退耦时期，即大爆炸后稍少于100万年时，是物质与辐射相互密切关联的最后时光。所以我们今天看到的宇宙背景实际上是宇宙那时候的风光。宇宙背景是一致的、各向同性的和均匀的这一事实表明，宇宙整体在 $t=0$ 后的700 000年时是一致的和均匀各向同性的。这是我们对大爆炸最靠近的直接观测。但不要忘记原初中微子。原则上它们是可以被探测的，按照标准模型的方程式，它们应该形成一个充满着今日宇宙的背景海，其温度是光子背景温度的70%，即约2K。它们的退耦是在 $t=0$ 后刚1秒时。如果这些中微子被探测到了，那就会是对标准模型准确性的最激动人心的证实，并能使我们看到宇宙年龄只有1秒时的那么遥远的景象。

在物质与辐射将近退耦时，整个宇宙的状况与太阳表面相似。它

很热，不透明，充满着黄色的光。随着物质与辐射的退耦，它突然变得透明，大约与此同时，辐射能量密度也降到了物质的等价密度以下。从大约 $t=10^6$ 年时起，宇宙变成由物质和引力主宰。可以借助于红移获得这发生在多久以前的一点大略印象。已知红移最大的天体是几个类星体，红移 z 在 3.5 到 4 之间。[6] 退耦时期和物质开始主宰宇宙的时期都大约对应 $z=1000$。宇宙微波背景中每个光子的波长，从它最后一次与物质相互作用时算起已被拉长了 1000 倍。

虽然初看起来辐射时期与最初 4 分钟相比似乎并无建树，但却有可能正是那时的不规则性后来增长而导致星系和星系团的出现。在辐射时期的末尾，当稳定原子刚刚形成时，宇宙中每升体积里大约有 1000 万个原子。今天平均说来在每 1000 升左右的空间才有 1 个原子。退耦时原子的数密度至少是今天星系中物质密度的 1000 倍，所以我们所知的星系显然应该是退耦后形成的。但是正在成为宇宙主宰的物质可能继承了在辐射时期的不规则性。某些地方的密度已经稍大于别处。到物质完全主宰一个透明、黑暗并正在冷却的宇宙的演化时，它已经聚成团块，而且由于自身的引力，这些团块不会像宇宙整体那样变得稀松。在这种有着上述平均密度的团块内，有些区域形成了气体云并开始碎裂和收缩，最后成为我们银河系和其他星系中的恒星。到宇宙已度过它现在年龄的一半时，银河系已经大体上是我们今天看到的这个样子；45 亿年前，太阳及其行星系统形成，形成它们的星际物质已经在许多恒星内部加工过又再加工过，因而含有丰富的重元素，当然还有那从大爆炸继承来的氢和氦。

在过去 100 亿年的绝大部分时间里，宇宙中的绝大部分物质是束缚在恒星和星系里，仅有的大尺度变化是星系团随宇宙膨胀的平静分离，以及背景辐射平静地不断红移和冷却。但是星系形成的详细过程

仍不清楚，天文学家们提出了一些相互竞争的理论，试图解释宇宙中的物质怎样聚集成我们今天看到的景象。粒子物理学的最新进展又一次提供了线索。的确，大多数在标准模型里尚未回答的问题答案，并不能在对今日宇宙的认识里找到，但是却能在对宇宙的第一个百分之一秒，也就是对标准模型所描述的大爆炸之前的认识里找到。

留下的问题

标准模型虽然在20世纪60年代和70年代取得了巨大成功，却也留下了一些未能回答的问题。那数量很少的物质(与光子数相比)是从哪里来的？为什么宇宙是这么特别的一致和均匀？为什么宇宙的密度是如此接近于使它平直？在从 $t=0$ 到第一毫秒末的间隔里发生了什么？最后，宇宙又是怎么来的——在创世时刻本身到底发生了什么？

我们又回到了形而上学的本质问题，曾被认为是超出科学范围的问题。但是现在科学已没有什么界限，而能向所有这些问题发起进攻，即使答案可能还不完整，或不能被完整地理解。感谢像剑桥的霍金这样的研究者，我们甚至有了向创世本身这一问题进攻的路线。对所有这些深刻奥秘的认识，都依赖于对在第一秒的一个很小部分里所发生事情的把握，那个时间是在由标准模型接着讲的故事之前，也就是说是在大爆炸本身之前。如马萨诸塞大学的宇宙学家哈里森所评论的，“第一个千分之一秒里宇宙史的内容比此后100亿年里的还要多。”但是要认识那一段宇宙史，那些已经很好地说明了宇宙的后继演化的物理定律就不够用了。需要加进来的是量子物理定律，即关于很小世界的物理定律，尤其是那些适用于极高能量密度的粒子和辐射的定则。所以在探索大爆炸之前的时期以前，必须先来了解那些量子

定则，领略一下它们的奇特。

注　释：

1. 数学家会说，严格地讲不可能让钟从 $t=0$ 开始，因为没有办法从奇点出来，也就没有什么 $t=0$ 的"时刻"，在奇点上时间概念是没有意义的。我对这些数学家的回答是，可以把 $t=0$ 设为尽可能地靠近奇点。这样做所产生的差异对目前的叙述是无关紧要的，但在引进量子物理效应并试图认识创世瞬间本身时就很重要，有关的研究结果将在本书第三篇中介绍。

2. 诺维科夫写于 1978 年的书《宇宙的演化》(*Evolution of the Universe*)对那个最初阶段作了略微详细和专门、但仍然很有可读性的讲述。如果把这两本书拿来一起读，就能很清楚地了解宇宙学家在 20 世纪 70 年代末对宇宙的认识。

3. 所以只有能量高于 1MeV 的光子才能生成一对正负电子。

4. 除了通过引力作用之外。中微子海所储藏的能量对整个宇宙的引力是有贡献的，已有人猜测中微子并非完全没有质量，而是可能有着几个 eV 的质量。果若如此，那么由于它们的数量巨大，其总质量就会成为宇宙总质量的主要部分，于是就会对宇宙的终极命运有深远影响，或许就保证了宇宙是闭合的，而不必再管氘丰度对核子物质密度所作的揭示。

5. 重核在早期的高密度下确实可以形成，但很快又分裂，从来就不能从火球中"冻结"出来。只有氢、氦、少量的氘和更少量的锂，加上中微子和其他非核子粒子，才能从大爆炸中出现。

6. 红移 z 是由光谱被拉伸的量来定义的。如果一个与光源相对静止的人测得的光谱中某一特征波长是 λ，那么另一个与光源有相对运动以致在他看来光源是在退行的人会发现那个特征波长变长了。两个波长之差记作 $\Delta\lambda$，则 $z=\Delta\lambda/\lambda$。例如，实验室里黄色钠 D 线的波长是 589 纳米。如果从一颗恒星或别的天体的光谱中找出的钠 D 线是在比如说 600 纳米的波长上，那么该天体的红移就是11/589，即 $z=0.018\,68$。对这种小红移，相应的退行速度就是光速 c 乘以 z，在现在这个例子里结果就是约 5000 千米/秒。对大于约0.4 的红移，就必须考虑相对论效应而用稍复杂一点的计算公式，即对退行速度为 v 的天体应有 $1+z=\sqrt{(c+v)/(c-v)}$。当红移为 2 时，这个公式给出退行速度为 $0.8c$，即光速的 80%。但不论红移多大，v 总不会超过 c。

插曲二　量子物理学

问题比答案更多——我发现得越多，我就知道得越少。

——纳什(Johnny Nash)

广义相对论是对大尺度时空已有的最好描述，单由它就足以得出，宇宙诞生于一种超密态，也就是确曾有过一个大爆炸。但是那产生广义相对论的科学革命只是20世纪初的两大革命之一。另一个突破则打开了对很小尺度(在原子之内)和很高能量物质的行为的新视野，在那种情况下物质和能量如爱因斯坦预言的那样变得可以相互转换。这就是量子物理学。当我们试图让宇宙真正起源的问题从形而上学领域进入科学领域时，单靠广义相对论是不足以胜任的，我们还需要量子物理学去描述很久以前发生在远小于1秒的时间和微小的空间里的极高能事件。所以，在探索回到比单由爱因斯坦理论所描述的大爆炸更早的时间之前，有必要介绍一下这个物理学史上最奇特和最有威力的理论的一些显著特征。在19世纪末，物理学家认为物质世界

是由很小很硬的粒子即原子和分子组成，它们相互作用而产生出在我们周围看到的各种有生命和无生命的物质。他们还有关于光如何以电磁波的形式传播的很好理论，这种波在许多方面与池塘里的水波或是通过空气振动来传递信息的声波相似。引力稍微难以理解一些。但总的说来，把世界划分为粒子和波看来是再清楚不过的，物理学已经接近于能解释一切。总之，理论物理的终结和所有重要难题的解决已经在望。

但是，就在物理学家准备去迎接这种舒适的可能性时，他们辛勤建造的纸牌大厦倒塌了。原来，光的行为有时候只有用粒子即光子的概念才能解释，而波的解释或模型只是在其他情况才保持适用。物理学家稍后又意识到，如果波有时表现得像粒子还不足以令人烦恼的话，那么粒子有时也可以表现得像波。与此同时，爱因斯坦正在用他的相对论推翻已经建立的关于时间、空间和引力本质的认识。当20世纪20年代末尘埃落定时，物理学家有了一幅世界的新图景，它与老的大不相同。这幅图景仍然是我们今天认识世界的基础。它告诉我们，没有单纯的粒子或波，在基础层次上讲只有波和粒子的混合物，偶尔被称作“波子”。它告诉我们，不可能以绝对的确定性来预言任何原子实验或宇宙中任何事件的结果，我们的世界是由可能性或者说概率来支配的。它还告诉我们，不可能同时知道一个微观物体的精确位置和它的精确动量(它正在向何处去)。

我已经在《薛定谔之猫探秘》一书中详细讲述了物理学家是怎样和为什么会得到这些惊人结论的。这里我只给出这幅世界新图景的轮廓，而不赘述有关的历史和实验详情。但量子物理学的基础是可靠的，它与爱因斯坦广义相对论一样，由实验和观测提供坚实基础和给予严格确认。它们一起作出了我们所能有的对宇宙及其中一切的最好

描述；而且正如在下文中将要看到的，在它们的领域重叠之处，它们结合起来去解释宇宙最先是怎么来的。

广义相对论涉及的对象是很大尺度的，即宇宙本身。量子物理学所涉及的则是很小尺度的，即原子或更小物体。两者的“重叠”是当宇宙在某种意义上只有原子核那么大，而其密度大于中子或质子的密度的时候。这就是为什么需要相对论和量子物理学相结合以解释宇宙的创生。这是本书第三篇的主题，在进入之前我们必须有所准备，也就是首先要有点量子物理学的知识。

光子

故事最好是从伟大的苏格兰物理学家麦克斯韦在 19 世纪第三个 1/4 期间的工作讲起。麦克斯韦 1831 年出生于爱丁堡，对物理学作出了许多贡献，其中最重要的无疑是他的电磁学理论。像许多同时代人和紧随而来的后继者一样，他也为通过一根导线的电流能产生磁场这一现象所强烈吸引，这种磁场的基本性质同磁体自身的磁场完全一样。例如，通电导线周围的磁场也能使其附近的小磁针偏转。而且，一个运动磁体经过一根导线附近时，还会在导线中激发电流。运动的电荷即电流能产生磁场，运动的磁体也能产生电流。电力和磁力曾经被认为是两种互不相干的现象，现在看来是同一个更大整体即电磁场的不同表现。

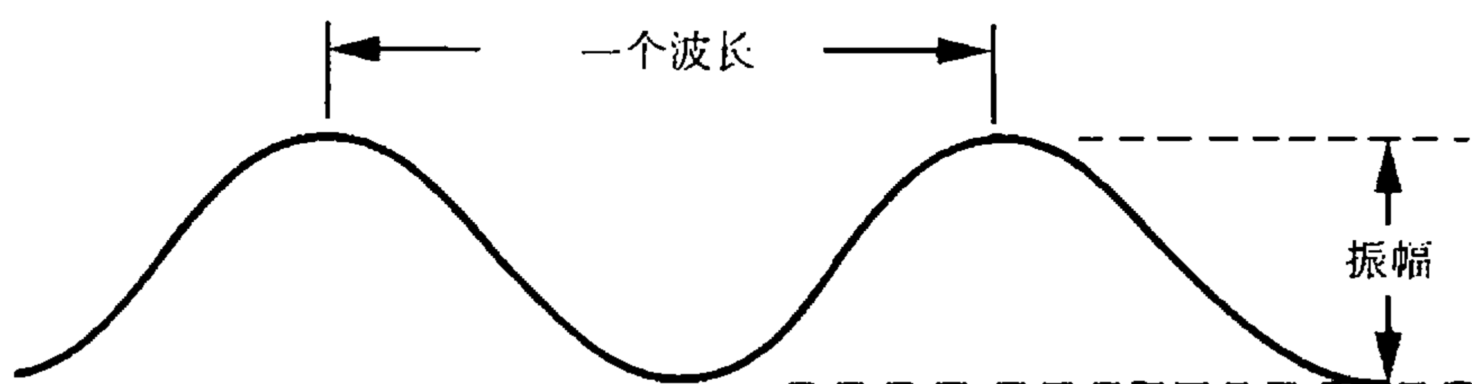
麦克斯韦试图写下一组方程来把物理学家已经观察到和测量过的所有电和磁的现象联系在一起。共有 4 个方程，一个是描述电流所产生的磁场；第二个是描述变化磁场所产生的电场；第三个给出电荷本身所产生的电场；第四个则给出对磁场本身的描述，包括磁极总是成对存在(一个南极一个北极)的奇特现象。但当麦克斯韦检查这些方程

时，发现它们在数学上有缺陷。为了弥补，他不得不在第一个方程里引入了另外一项，那一项等于是描述没有电流时磁场如何能由变化的电场产生。

那时候从未有人观察到这种现象。但是一旦麦克斯韦把方程组改造成了那种最优美的形式，那个附加项的道理也很快就清楚了。物理学家知道电容器，那是两块隔着小空隙的平直金属板，两板之间可以有电势差。如果把一块板与电池的正极相接而另一块板与负极相接，那么前一块板就会带上正电荷而另一块则带负电荷，两板之间的空隙就是一个强电场区域，但是没有电流流过空隙，也没有磁场。麦克斯韦的新数学项的功能之一，就是描述当电容器板刚刚与电池接上时在两板之间会发生什么。当板上电荷积累时，板间空隙里就有一个迅速变化的电场，按照方程式这就会产生磁场。麦克斯韦简单地把一根小磁针放到板间空隙里，并观察它如何在板与电池连接时偏转，从而很快就证实他的方程是正确的。像所有最好的科学理论一样，新电磁理论也成功地预言了实验的结果。

下面才是真正惊人的发现。麦克斯韦意识到，如果变化的电场能产生变化的磁场，变化的磁场也能产生变化的电场，那么这两者就能作为一个统一的电磁场的两个成分而很好地相依相存，并不需要任何电流或磁体。按照他的方程式，这样一个由电生磁和磁生电而自行增援的电磁场，一旦被首次产生就很容易自己在空中前进。由方程所预言的连续变化的电磁场是以波的形式运动，有着一定的速度即每秒 300 000 千米。这正是光速。麦克斯韦的电磁学方程预言了以光速运动的电磁波的存在，却没有使他进一步认识到光必定是一种电磁波。

那时已经有很确定的证据表明光是波动的一种形式，麦克斯韦的发现合乎 19 世纪科学的主流，因而大受欢迎。关于光的波动本质的

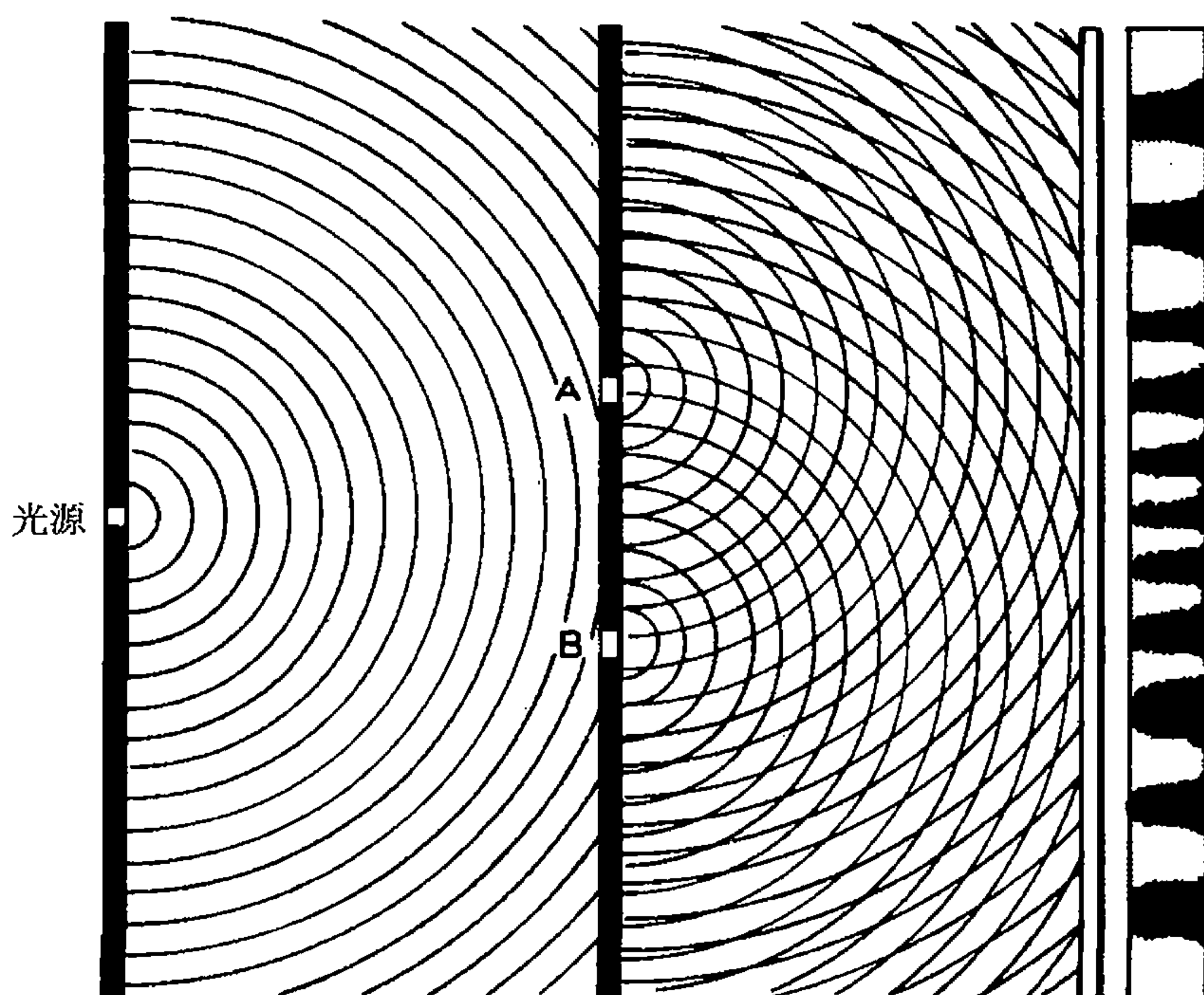


插曲图 2.1　波

最好证据是其“干涉”现象，像池塘里两束水波的干扰所形成的起伏一样，光的干涉也生成明暗相间的条纹，这是不能作其他解释的。托马斯·杨(Thomas Young)，一位 1773 年生于萨默塞特的英国物理学家和内科医生，于 19 世纪初得出了关键的实验证据。他让一束单纯一种颜色的光(单色光)穿过一块屏板上的两个窄缝而形成两股“波浪”，从而在下一块屏板上得到典型的干涉图案。这个工作有力地否定了由牛顿提出的光以很微小粒子的形式出现的老概念。

麦克斯韦和杨两人的工作结合起来，似乎对光的本质有了透彻的认识。由干涉实验可以测量光的波长，即一个波峰到下一个波峰的距离，结果大约是 10^{-7} 米；不同颜色的光有着不同的波长，红光的波长要比蓝光长几倍。麦克斯韦推断，在远超出可见光谱的整个波长范围都必定有电磁辐射，有的辐射波长比可见光的短得多，有的则长得多。一位德国的先驱者赫兹(Heinrich Hertz)在 19 世纪即将结束时造出了波长为几米的射电波，从而证实了麦克斯韦的预言。

从射电波到可见光再到 X 射线的全部电磁波谱都服从麦克斯韦方程。这些描述电磁辐射如何作为波来传播的方程，是诸如电视机和收音机之类的日用电器的设计基础。它们也是对红移作出宇宙学解释的基础，光是一种波的认识也的确已是众所周知，所以我在前面讲红移效应时已经用了波长概念，而没有先介绍杨和麦克斯韦的工作。但是从 20 世纪初起人们又转而认识到，原来牛顿也是正确的。光和所



插曲图 2.2 两个波相遇时会互相干涉。托马斯 · 杨用这一现象证明光是一种波。从一个光源来的光穿过屏板上的两个孔 A 和 B 而成为两束波。在右方屏板上波互相抵消的地方就留下暗纹，波互相加强的地方则出现亮带。

有的电磁辐射也都能表述为现在被称为光子的粒子。有些情况下光的行为只能用光子的概念来解释，这是爱因斯坦在 1905 年指出的。

关于光的粒子本质的第一个启示是在 1900 年出现的，那时普朗克(Max Planck)，一位 1858 年出生于基尔的德国物理学家，发现自己不得不在描述光或其他电磁辐射如何由高温物体发射出来的方程中引入不连续的光包的概念。这是 19 世纪 90 年代物理学家面前的一个重大难题。他们猜想，光是由物体内带电粒子的振荡产生的，原子本身也参与振荡，这种振荡正是麦克斯韦方程所描述的波出发和前进所需要的推动。他们又由观测和实验知道，物体产生的辐射的性质取决于其温度。我们从日常经验中也知道这一点，白热金属块(例如火钳)的温度比红热金属块高，一块太冷而不足以发出可见光的金属仍会太烫而不敢用手去碰。任何一个物体都在很宽的波长范围内发射电磁波，

但是辐射强度的峰是在一个由物质温度决定的波长上，温度越高，这个峰值波长就越短。整个辐射谱的形状都一样，而峰的位置则精确地反映出辐射温度，这种谱就是由一个“完美”辐射体产生的著名的“黑体”谱，宇宙背景辐射就具有这种黑体谱，相应的温度略低于3K，但在普朗克登台之前，没有人能这样地操作电磁方程，以从中得出黑体谱的性质。

普朗克发现，解释黑体谱观测性质的唯一途径是，光(我现在是指任何一种电磁辐射)是由原子内的振荡电荷以一种很小的能量包的方式发射出来。[1] 这意味着原子对光的吸收也只能这样一份一份地进行。普朗克把这个概念用辐射频率 ν 来表达。频率可以理解为每秒钟经过一个固定点的波峰的个数，对波长为 10^{-7} 米、速度为每秒300 000千米的光来说，频率就是每秒 3×10^{15}，或曰 3×10^{15} 赫，以纪念赫兹这位无线电的先驱者。普朗克发现，观测的黑体谱可以这样来得到解释，即对每一个频率的光都有一份特定的能量，它等于频率乘以一个基本常数，他把该常数记作 h 。这个能量即 $E=h\nu$ ，是任何原子以该频率所发射或吸收的能量的最小份额，原子所发射或吸收的能量只能是这个基本单元的精确整数倍(1，2，3，4，… n ，…)。

普朗克并没有提出光的能量只能以这种 $h\nu$ 小包的方式存在，他认为对发射或吸收的限制只是由物体内振荡电荷的某种性质所致。但他算出了那个对所有辐射都相同的 h 的值。它现在被称作普朗克常量，其值的确很小，为 6.6×10^{-34} 焦/秒。即使对频率为 10^{15} 赫的光，能量的基本单位也只有约 10^{-18} 焦，而一个电灯泡亮 1 分钟的能量就有 6000 焦。正因为 h 很小，灯泡所发射的光才看似连续，而事实上可见光是由亿万个小能量包组成的。

普朗克的建议受到了欢迎，但不是很令人满意。它看来是解释了

黑体谱，但却是通过一种数学手法、一种技巧。那时在瑞士专利局的办公桌前一位几乎无人知晓的物理学家，也就是爱因斯坦，使那个数学技巧成为了一种物理真实。他证明，如果那些小能量包是一种真实存在，也就是光只能由这种能量为 $h\nu$ 的小单元组成，那么当时的另一个重大难题也就能迎刃而解。爱因斯坦对光的本质的探索，对其所以然给出了一幅清楚得多的物理图像。

光电效应是当光照射到真空中的金属表面上时发生的。光把金属中的电子撞了出来，电子能被探测到，其携带的能量也能被测定。这个现象是 1899 年由匈牙利人勒纳(Phillip Lenard)发现的。发现光的能量能使金属中的电子跳出来倒不是很令人惊讶，惊人的是光能量与电子能量之间的关联。勒纳用的是单色光，也就是说所有的波都有同一的频率。一束明亮的光携有的能量显然比一束暗光多，所以似乎应该预期，当亮光照到金属表面时，撞出来的电子会具有较高的能量。但勒纳的实验结果却是，只要用的是同样频率的光，那么不论光是多亮，出来的电子相互并无差异，每个电子都有着同样大小的能量。

当勒纳把光源移近金属，因而金属表面更明亮时，的确有更多的电子由光电效应产生出来，与从光源得到更多的能量相应。但是那些电子中的每一个所具有的能量是一样的，而且也与暗弱光源情况下每个光电电子的能量相同，尽管在后一种情况跳出来的电子较少。另一方面，如果他改用较高频率(即较短波长)的光，那么即使光源暗弱，所产生的电子也有较高的能量。这些电子的能量仍然都一样，但都比由较低频率的光所产生的电子能量高。所有这一切的原由事后看来都很简单，但是爱因斯坦的建议在当时的确是革命性的。他建议(当然还给出了作为支持的数学计算)，一束频率为 ν 的光是一股由粒子(我们现在称为光子)组成的流，每个粒子都有能量 $h\nu$ 。当一个光子以适

当方式撞击金属中的一个原子时，原子中的一个电子就会被射出来。于是每个射出电子所带有的能量，就应该是光子所提供的能量 $h\nu$ 与使电子脱离原子所需的能量之差。光越明亮，就有越多的光子，所以就会有越多的电子产生出来。但是每个光子的能量是一样的。使单个光电子能量增大的唯一办法，就是增大撞击金属的光子的能量，而要增大光子能量就只有使 ν 增大。

这些意见远不是当时的物理学家所普遍欢迎的。大家都知道，杨氏双缝实验和麦克斯韦方程组已经无可置疑地确定了光是一种电磁波。一个新来的看来并不真正懂得什么物理学的轻率家伙，竟敢要复活牛顿那个已经错了的光微粒概念。有一位美国实验家密立根(Robert Millikan)非常愤慨，竟用了10年时间去做实验，意在证明爱因斯坦的假说错误。但他恰恰成功地证明了爱因斯坦的正确。他得到了 h 的一个很精确的测量值，从而帮助了普朗克在1918年获得诺贝尔物理学奖，也帮助了爱因斯坦在1921年以对光电效应的解释而获得该奖，最后他自己也在1923年获奖。所有这些奖无疑都是应得的，令人不解的是爱因斯坦没有以他的广义相对论而第二次获得诺贝尔奖。

在这些量子论的先驱获得荣誉时，普朗克引入原子物理学的量子 $h\nu$ 已经帮助以丹麦人玻尔(Niels Bohr)为首的其他一些物理学家建立起第一个令人满意的原子模型。这个模型的大意是，原子中有一个很小的带正电荷的核，更小的带负电荷的电子在围绕核的轨道上运转，多少有些像行星环绕着太阳。电子的轨道是分离的，轨道间的间隔与能量的基本量子相应，电子能从一个轨道跳到另一个，但不能存在于两轨道之间。如果它从高能轨道跳到低能轨道，它就会发射出一个能量为 $h\nu$ 的光子；而要从低能跳到高能就必须吸收一个光子即 $h\nu$ 的能

量。由于不能有半个光子，电子也就不可能跳到两个容许轨道之间。

这个模型远非完整，但却使物理学家能把握原子中电子的行为方式，从而去解释原子光谱中的明线和暗线。电子在能量阶梯上跳下一级或几级时发射光子就表现为明线；电子跃上阶梯时吸收确定频率的光子，光谱中留下的“空缺”就是暗线。但在 20 世纪 20 年代初还有许多尚未解决的问题。首先，电磁辐射的理论不是一个而是两个。有时光和 X 射线必须用麦克斯韦的波方程来描述，但有时又不得不用爱因斯坦的光子，有时还得用这两者的混合，如普朗克计算黑体谱时那样。而最紧迫的问题是，决定原子中的哪些能级能被电子占有的是什么，或者说能量阶梯上横档的数目和其间的间隔是怎么确定的？答案不是来自如何使光的理论合理化并回归到 19 世纪“经典”物理学有条不紊的逻辑性，而是来自将认识波的这场革命扩展到粒子世界。尤其是，如德布罗意(Louis de Broglie)在 1924 年向大吃一惊的物理界提议的，既然波可以表现得像粒子，为什么被认为是粒子的电子就不能表现得像波呢？

电子

德布罗意生于 1892 年，在我于 1985 年春写这句话时他还活着。*他是一个法国贵族的小儿子，后来从哥哥那里继承了家族封号而成为德布罗意公爵。他哥哥莫里斯(Maurice)是 X 射线光谱学的创立者之一，他正是从哥哥那里得知量子革命并为其深深吸引的。他在 1924 年提交给巴黎大学的博士论文里提出的思想简单明确，又以完善的数学分析为支持。我这里只简单勾画德布罗意关于物质本质的物

---

* 德布罗意卒于 1987 年。——译者

理图像，有关的数学则请到别处查找。

爱因斯坦已经给出了物质粒子的能量公式 $E=mc^2$；普朗克在爱因斯坦的一点小帮助下也对光子得到了一个类似公式 $E=h\nu$ 。光子虽然没有质量，却确有动量，否则它就不可能把电子撞出金属表面。普通粒子的动量等于其质量 m 乘以速度 v 。一个迅速运动的轻物体，能携带与一个运动缓慢的重物体一样多的能量，并作出同样强的撞击。想想一粒子弹的冲击并与一只垒球作比较就能明白。当然，最强的撞击是来自质量大而又运动得很快的物体。光子的速度是 c 。如果把爱因斯坦公式中的因子 mc 用代表动量的字母 p 代替，就得到一个新公式 $E=pc$ ，它能同样适用于普通物质和光子这两者。

德布罗意于是把这个式子同普朗克公式合在一起，$E=pc=h\nu$ 。这样就对光子得出 $p=h\nu/c$ 。但波的速度 c 除以频率 ν 就正是波长。所以德布罗意版本的公式说的是，光子的动量等于普朗克常量除以其波长。这就把光子的粒子性(动量)与其波动特征(波长)用普朗克常量直接联系起来了。但是德布罗意说，为什么就此停住呢？电子有动量，普朗克常量的值也是知道的，由上面的公式就得出其波长是等于 h 除以其动量。换句话说，任何粒子，比如电子，也是波，其波长仅由其动量和普朗克常量决定。日常物体的质量和动量与 h 相比太大，h 除以动量几乎等于零，故日常物体的波动性可以忽略。但是电子的质量只有 9×10^{-28}克，其波动特征就变得很重要了。

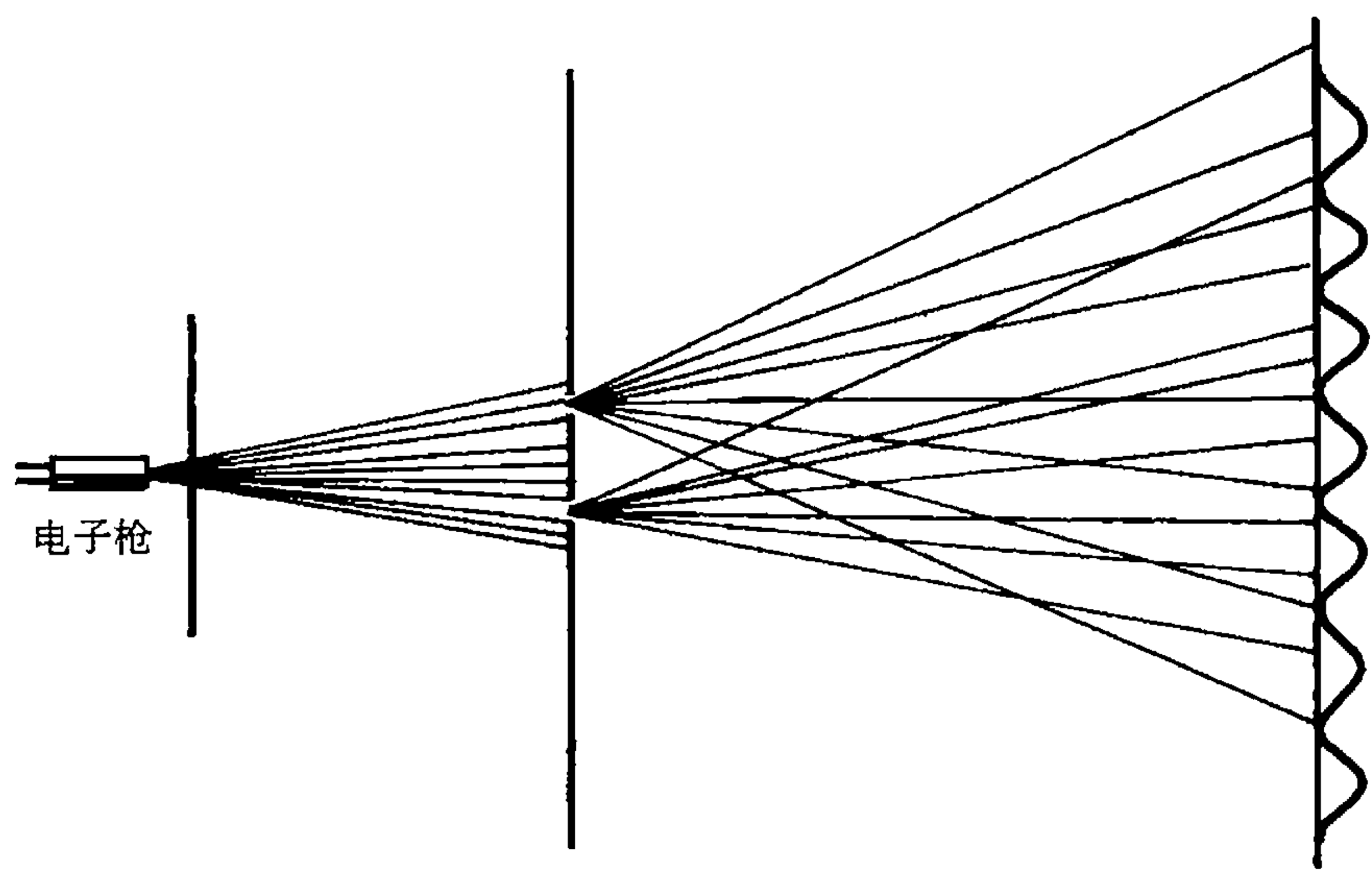
德布罗意告诉考官们，这个奇怪的公式有着物理真实性，可以用实验来测量电子的波长。考官们并不重视这个主意，把他论文的这一部分看作只是一个聪明的数学游戏，而不认为有什么实际价值。他还是得到了博士学位，他的导师朗之万(Paul Langevin)把他的论文寄了一份给爱因斯坦。爱因斯坦立即看出了它的含义和价值，并转告给了

其他研究者。只过了几年，美国和英国的研究组就都真的测出了电子的波长，所用的实验是让一束电子被晶体中排列整齐的原子所散射。对电子来说，这就相当于是杨氏双缝实验，而且同样是结论性的。只有波才能相互作用而产生干涉图样，在适当条件下电子也能如此。与此同时，奥地利的薛定谔(Erwin Schrödinger)建立了电子的波动方程，与麦克斯韦关于光的方程相当，它是发展出一个自洽的原子模型的关键之一。德布罗意被及时地授予了1929年的诺贝尔物理学奖。到那时已经清楚，所有的“波”都必须被看作“粒子”，所有的“粒子”也必须被看作“波”。对于日常生活中大到可以看见的物体，或是海洋和池塘里的波浪，这不会引起混淆。但是对于认识原子和亚原子现象来说这却是至关重要的。20世纪20年代后期的物理学家很高兴终于有了一个关于原子的自洽理论，即使他们必须付出代价，包括了关于波粒二象性的奇怪思想。但是，真实原子的这种奇怪特征可以说只是后来被打开的量子世界怪事里最小的一件。

中心奥秘

有这么一个实验，它包含了量子物理学的实质，直指其中心奥秘。这是杨用以证明光是一种波的双缝实验的现代样式。这个实验可以用光或电子或其他粒子如质子来做，也并非一定是刻板地要用两道缝而可以是某种等效装置，比如所谓的衍射光栅，或是晶体中使入射的X射线或电子弹开的排列整齐的原子。但是为了以最纯粹的形式来描述这一中心奥秘，我这里只讲电子和恰恰是两道缝的情况。我所讲的电子的一切行为都已由此类实验核查和证实，而且，不仅对电子，对光子和其他“粒子”也都已做过。其中没有一样只是来自方程式的古怪数学猜想，而是全都测试过、验证过的，是真实可靠的。

实验的理想化形式很容易描述。一个电子源(或叫电子“枪”，就像电视机里的显像管)，一个开有两个小孔的屏板(孔必须小到可与电子的波长相比，这就是晶体中原子之间的空隙正合适的缘故)，还有一个探测器。探测器可以是一个屏幕，就像电视机的屏幕，电子打上去会闪光。重要的是有一种手段来记录电子在什么时间打到了探测器上的什么位置，并且把到达探测屏幕上每个位置的电子数目显示出来。当波通过这样一个装置时，第一块屏板上的两个孔就都成了波源，每一个孔出来的波都成半球形铺开并与另一个孔的波同步。在两个波相加的地方，就会出现大的振荡；而在它们相抵消的地方，就根本探测不到任何振荡。所以这个实验用光波来做时就会是探测屏幕上出现明暗相间的条纹。



插曲图 2.3　用光做的杨氏双缝实验(插曲图 2.2)也同样可以用电子来做。实验结果也完全一样，即证明了电子是波。而别的实验也同样可靠地“证明”光和电子以粒子的形式存在。这种波粒二象性是量子力学的中心奥秘。

当然，如果第一块屏板上只有一个孔，那么在探测屏幕上就简单地只是一个光斑(或光带)，其中心最亮，周围则逐渐暗下去。

若是让一束电子从屏板上的一个孔(或一条缝)通过，也会发生同

样的事情。探测屏幕上正对着孔的位置上记录到绝大多数闪光，在这个最强位置周围则有少数闪光散布。在双孔装置里，可以先后遮住一个孔只留下另一个孔来作验证。一个孔总是使屏幕上出现一个亮斑，斑的周围逐渐暗淡。但当两个孔都打开时，屏幕上就出现清晰的干涉图样。标记单个电子到达的闪光形成由暗带隔开的明亮条纹。这只能由电子的波动性质来解释。穿过两个孔的电子波互相干涉，在一些地方抵消而在另一些地方加强，正如光波一样。

到目前为止，一切正常。当然是有点怪，电子在通过实验装置时的行为像波，然后又突然聚合成小而硬的颗粒在探测屏幕上撞出闪光，但若把粒子和波的概念结合起来，我们至少还能使自己觉得多少能明白是怎么回事。水波毕竟也是由大量运动着的小粒子(水分子)组成的。既然有成千上万的电子穿过那两个孔，那么它们像波那样运动而又保持其为小粒子，似乎并不很令人惊讶。如果一次只射出一个电子，那么似乎很有理由预期它将是穿过这一个孔或那一个孔。按照日常推理，屏幕上的干涉图样应该就是由许多电子在同一时间造成的。

如果真是一次射出一个电子，究竟会怎样呢？屏幕上若只有一次闪光，那么显然不足以说明电子的行为。但是可以一次又一次地重复这种点射，观察所有的闪光并注意它们在屏幕上的位置。这样就会发现，闪光慢慢地构造出原来见过的干涉图样。单独通过该装置的每一电子都不知怎么地像个波，自己干涉并沿自己的路径落到图样上的某一亮区。这结果就像是在不同时间通过装置的所有电子在互相干涉，或者是互相都有“记忆”，以形成干涉图样。

看来好像是每个电子都穿过两个孔。这简直是疯了！但我们可以发明另外一种探测器，以记下每个电子是通过哪个孔，并重做实验来看看究竟是怎么一回事。结果并不是两个孔边的探测器都报告通过了

同一个电子(或半个电子)，而是有时电子通过这个孔，有时通过另一个孔。如果先关上一个孔让大量电子从另一个孔通过，然后再将两孔互换，那么屏幕上就简单地只是两块亮斑而没有任何干涉的迹象。

这真是太奇怪了。当我们试图探测一个电子时，它的反应像个粒子。但当我们不去看它时，它的行为又像是波。当我们要看它是穿过哪一个孔时，它只穿过一个孔而不顾另一个孔的存在。但当我们不去跟踪它的路径时，它不知怎么又能同时“知道”两个孔并且好像是从两个孔穿过来的。

量子物理学有很好的术语来表述这一切。它说，有某种类型的波与电子相联系，它叫做“波函数”，而且原则上可以展布到充满整个宇宙。薛定谔方程描述的就是这些波函数和它们的相互作用。波函数最强的区域，就是日常所说的电子位置。从这个区域往外，波函数就会减弱，但即使是在远离电子“位置”的地方也依然存在。薛定谔方程能很好地预测像电子这样的粒子在不同情况下的行为，包括这些粒子或波函数通过两个孔时将如何互相干涉。当我们看一个电子也就是说用粒子探测器去测它时，波函数就“坍缩”了。在那个瞬间，电子的位置是能在基本定律允许的精度范围内确定的。但一旦我们停止去看，波函数就再次散开并与其他电子的波函数相干涉，而且在适当条件下也可以自己与自己干涉。[2]

所有这一切都是可精确地用数量来表示的，因而可以用来计算电子是怎样置于原子之中，原子又如何结合成分子，以及其他许多。“波函数的坍缩”这句难懂的话(它在量子理论中有着精确的数学意义)等于是说，我们仅仅是在真正去看一样东西的时候才能知道它在何处，一眨眼它就跑了。粒子的行为取决于我们是否在看它们。如果我们盯着两个孔看电子如何通过，电子的行为就与我们不去看时不

同。在量子物理学里，观测者也是实验的一个必要组成部分，他或她选择看什么会对发生什么起决定性的作用。

所有这些的含义的确非常深奥。比如，我们不再能说一个电子可以作为一个物体从实验装置的一端出发，沿着确定的路径或轨道到达另一端。那个连续的“轨道”概念是经典的牛顿理论遗留下来的，现在必须抛弃。量子物理学讲的是“事件”，事件可以按一定的时间顺序发生，但绝不能从中得知参与事件的粒子不被观测时在干什么。我们所能说出的一切是，我们观测到一个电子出发(事件 1)，又观测到一个电子到达(事件 2)。我们根本说不出电子在这二者之间干了什么，我们甚至不能说这两个事件记录的是同一个电子。将两个电子一起射出，过了一会儿有两个电子到达了探测屏幕，但是无法区分哪个是哪个。

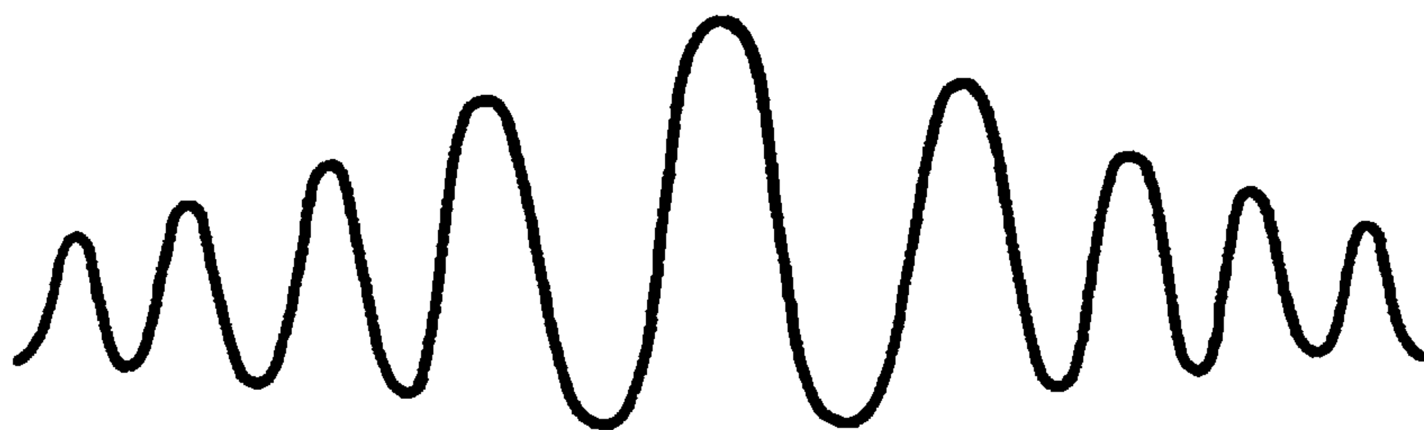
日常世界的有形物体，诸如夹纸用的回形针，似乎也互相不可区分，但电子的互相不可区分是在深刻得多的意义上。原子中的电子并不是什么各有自己确定轨道的清晰物理实体。我们所能说的只是，某一特定种类的原子的行为，像是与 8 个或 10 个或别的什么数目的电子波函数相关联。如果我们做一个实验去戳一个原子(比如说像光电效应实验那样用光子去轰击)，那么电子波函数中的一个或多个会改变，以至于很可能我们将在原子外面探测到一个电子，这小粒子像是被射出来的。只有我们观测到的才是真实的，其他一切都是推测，都是我们在脑子里构造的假想模型，我们用这些模型和方程式给出一幅正在发生着的事情的图像。

哪个更真实，是粒子还是波？这取决于涉及的是什么问题。无论物理学家多么善于提出问题，却从来不能绝对肯定将会得到什么回答。

偶然性与不确定性

粒子是有明确定义的。它存在于空间的一个点，占有一个很小的体积，具有我们日常经验意义上的可触知的真实性。波则几乎正相反。一个单纯的波可以永久延伸，根本就不能说它存在于一个点。它有动量，也就是说有一个明确的方向。但是即使在想象中你都不能用手指去碰它并使它停下来让你看看。亚原子世界的这两个方面是怎么调和的呢?

要用粒子概念来表述，比如一个光子或一个电子，波就必须以某种方式予以限制。数学家知道该怎么做。限制一个波的办法是降低其单纯性。设想有一个频率范围并一起运动的一组波，而不是只有一个固定频率的单波。在有的地方，一个波的峰会和其他波的峰合起来而形成一个更强的波；在别的地方，一个波的峰正好遇上别的波的谷，于是就互相抵消。数学家用一种叫做傅里叶分析的方法，能够描述一组波的这样一种组合，即除了一个很小很确定的空间范围外这些波到处都几乎完全抵消。这种组合称为波包。原则上讲，只要波包里放进足够多的不同的波，就能使波包要多小就有多小。数学家常用希腊字母 Δ 来表示小量，于是波包的长度就可以记作 Δx 。各有其单一频率的许多波组合的结果，就可以把波包限制到电子那么大的尺度。



插曲图 2.4　波包是只覆盖--个确定空间区域的一组波。

但也付出了代价。如德布罗意所指出的，波的动量是 h 除以波长。一个单纯的波有单一的波长从而有单一的动量。但是引入了混合的波长或频率也就引入了混合的动量。参与混合的波越多(即波包越小)，波包的“动量”的精确度就越差。我们所能说的只是，动量是在一个误差为 Δp 的范围内。 Δx 是波包位置的不确定性的量度，我们只知道它是在 Δx 范围内的什么地方。 Δp 则是波包动量的不确定性。我们能大略知道波包向何处去，但只能精确到 Δp 。数学上很容易证明，绝不可能把 Δx 或 Δp 减小到零，实际上这两个不确定量的乘积 $\Delta x \Delta p$ 总是大于普朗克常量 h 除以 2π。这个稍有不同的常量被记作 $\hbar$ ，要么被称作“h 杠”，要么就也称作“普朗克常量”，尽管它实际上是 $h/2\pi$ 。位置与动量的这种关系叫做海森伯不确定关系，以纪念诺贝尔奖获得者德国物理学家海森伯(Werner Heisenberg)，他是20 世纪 20 年代第一个完整形式的量子力学的创立者之一。[3] 该关系式就是：

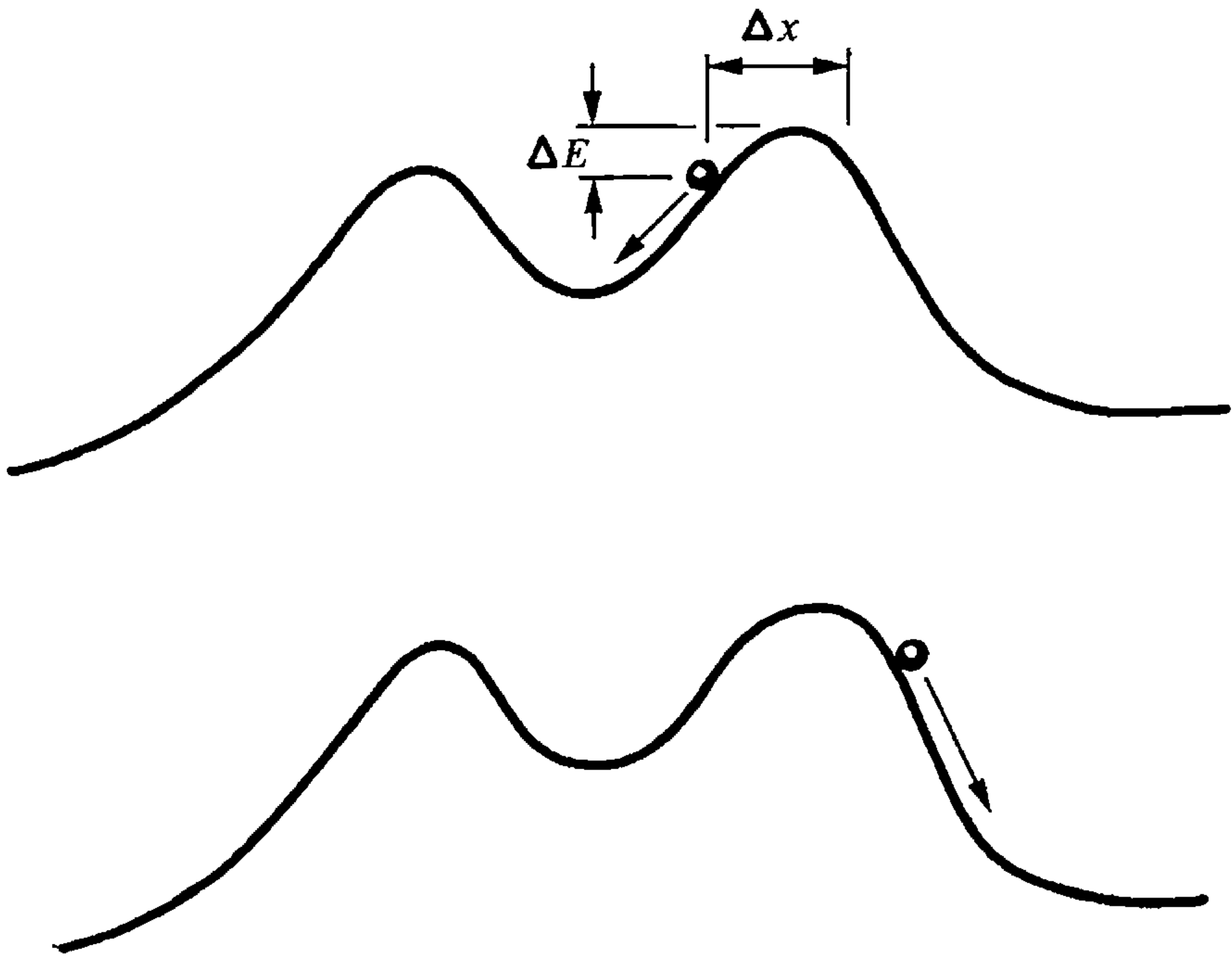
$$\Delta x \Delta p > \hbar$$

这个关系式不只是一种怪诞的数学技巧，对此怎么强调也不会过分。关于波粒二象性的经验证据表明，原则上不可能同时绝对精确地测量粒子的位置和动量。如果能精确测出一个电子在什么地方，即 Δx 为零，那么 Δp 就将成为无穷大，因而根本就不知道那个电子下一时刻会出现在宇宙中什么地方。不确定性并不限于指我们对电子的认识。它始终成立，是电子也是其他粒子和波的真正本质。粒子自身并不绝对精确地“知道”它现在何处并且下一步将去何处。在量子物理学里不确定性的概念与偶然性的概念有着本质的联系。我们不能肯定一个粒子在哪里，也不能肯定它向何处去，所以如果它出现在不是我们预期的地方，我们也不必太觉惊讶。

位置和动量并不是粒子以这种方式相联系的仅有性质。还有其他的所谓“共轭变量”对也类似地由波动方程联系着，其中最重要的是能量 E 和时间 t 。严格的数学分析表明，在亚原子过程中能量也有一种内在的不确定性。如果能量从一个粒子转移到了另一个，而且转移是在一定时间内发生(这是必然的，因为没有任何过程能比光更快)，那么能量不确定性 ΔE 与时间不确定性 Δt 的乘积也大于 $\hbar$ ：

$$\Delta E \Delta t > \hbar$$

对一段足够短的时间而言，粒子和它的邻近环境，或甚至是整个宇宙，都不清楚粒子具有多大能量。那个使 α 粒子逃出原子核的奇怪的“隧道效应”，生动地描绘了亚原子世界里不确定性的威力。伽莫夫当然是用完整的量子物理方程来“正经”地解释了 α 粒子发射，不过我们这里也不难用通俗的语言来理解。



插曲图 2.5 量子力学的不确定性解释了 α 粒子如何逃出原子核(图 4.6)。粒子能够短暂地“借到”能量 ΔE 从而“翻过山峰”；或者用另一种观点，它能开辟一条隧道穿过山峰，因为它的位置在 Δx 范围内是不确定的。

记住 α 粒子本是在原子核内，可以想象它是恰在火山口内侧一点点。如果它是在火山口外一点，它就会“滚开”即被电斥力弹出。从核内到核外的“距离”是 Δx 。一个与核相连的 α 粒子也与核一样有很确定的动量。但这就意味着它的位置不确定。虽然它与核相连，若按日常意义上的“内”来说，它不一定在核内。按照不确定性，就有这样一种有限的、可以精确计算出来的可能性，即粒子其实是在核外的可能性。瞧！有些粒子的确发现自己是在核外，于是就射出，就像它们由“隧道”穿过了隔着的山峰。这也正像你在杯子里放进一些骰子，再把杯子摇得格格作响，突然有一个骰子出现在杯子外面，并沿着桌面滚开。假如普朗克常量足够大，日常生活中骰子的行为就真如 α 粒子一样了。

也可以用能量概念来考虑。粒子需要更多的能量才能攀越势垒。它需要的这份额外能量是 ΔE 。在一段足够短的时间 Δt 内，它可以如按所有的物理定律所知(或者说所关心)的获得这份额外能量。这样它就也能出来并跑开。在时间 Δt 以后它必须“归还”从不确定性借来的能量，但这并不要紧，因为到那时它已经沿势垒的另一侧逃下山去了。

如果这些还不足以使你的脑子转弯，那么量子力学还有更多的妙招。到目前为止所讲的主要是把粒子看作波的混合。当波，尤其是光，必须作为粒子来处理时又如何呢？物理学家所遇到的此类问题可以由一个例子来集中体现。有一种器材叫偏振片，它只允许在一定方向上振荡的光波通过。有的太阳镜就是这样。由于阳光中包含有在所有方向上振荡的波，只让其中一部分通过眼镜就能减弱光强。“偏振”光就是经过了这种过滤器的光，剩下来的波全都在偏振片所规定的一个平面内上下振荡。如果这种波遇上另一个偏振片，它容许的偏

振面正好与波的偏振面垂直，那就没有光能再通过。但若第二个偏振片的偏振面与光波偏振面的交角小于 90°，那么有一部分光能通过。能通过多少取决于交角；若交角是 45°，那就正好有一半光通过第二个偏振片，偏振面与通过之前相比也转动了 45°。

麦克斯韦方程组能够解释这一切。一个成 45° 角的偏振片能去掉波的一半能量，一个成直角的偏振片就去掉了全部能量。但是对于来到偏振片上的单个光子来说发生了什么呢？不可能把光子砍下一半，它是能量的基本单位。所以在一束光遇到成 45° 角的偏振片时，只能是一半光子通过而另一半不通过。但光子是怎么被挑选的呢？按照概率的统计规则是偶然的、随意的。当一个光子到达偏振片时，它通过和被挡住的可能性精确地是 50∶50。这个比例可以随偏振片的角度改变，但原理是一样的。偏振片对单个光子的挑选纯粹是偶然的。这个例子就是整个量子世界所发生的事情的一个简单示范。每次当亚原子粒子参加相互作用时，其结果都有偶然性。某一个特定结果的可能性也许会很大，也可能并不比掷硬币时的 50∶50 更好。这些可能性全都由量子物理定律清楚地、精确地规定，在量子世界里是没有确定性这种事的。

正是这一点使得爱因斯坦拒绝接受整个量子理论，他的名言是“我不相信上帝在掷骰子”。但是所有证据都表明上帝确实在掷骰子。所有的实验都证实了量子解释的精确性。例如，我们若做一个实验，其中要求测量一个电子的位置，我们就不可能确定地知道它后来会怎样。在这种简单情况，我们能说的或许是，有一定的概率在下一时刻我们会看到电子已经从 A 点运动到了 B 点，还有一个不同的概率它得在 C 点，如此等等。原则上所有的概率都可以计算出来，电子从 A 到 B 的概率可以是远大于 99%，但绝不会是完全确定。有一天

当你重做那个电子从 A 到 B 的熟悉的老实验时，得到的结果却可以纯粹偶然地是到了 C 点或 D 点或 Z 点。

在日常世界里，完全是所涉及的粒子数量把我们从这种稀奇古怪的概率里拯救出来。亿万个电子在线路里运动，才使我的电脑工作。在系统中的某处可能真有几个电子走错了路，从 A 到了 D 而不是从 A 到 B。但对我所关心的事来说，绝大多数电子是在做它们应该做的。除非你是个有点哲学味道的人，你在自己的生活里不必太担心量子理论中的这种不确定性和偶然性。即使你像我一样真有那么点形而上学味道，你仍然可以使用电脑，不必真的害怕所有电子都会突然不服从指挥。但是在涉及亚原子世界和宇宙的诞生时，量子物理学的奇异规则就变得特别重要。我们还需要量子食谱里的一个基本概念和几样珍品，最后就能准备就绪去进攻那些难题。

路径积分与多种世界

量子力学与经典的牛顿力学之间的基本差别，在我们仔细察看一个粒子(如电子)从一个点 A 运动到另一个点 B 的方式时可以清楚地显示出来。按照经典的观点，粒子在 A 点有一个朝着确定方向的确定速度。它在外力作用下沿一条可以精确确定的路径运动，经过或终止于 B 点。量子力学的观点则不同。我们即使想在原则上同时知道粒子的位置和动量也不可能。对于粒子向何处去有一种内在的不确定性，如果粒子从 A 点出发，后来在 B 点被探测到了，那么除非我们沿着它的路径全程注视，否则就不可能确切知道它是怎么从 A 到 B 的。

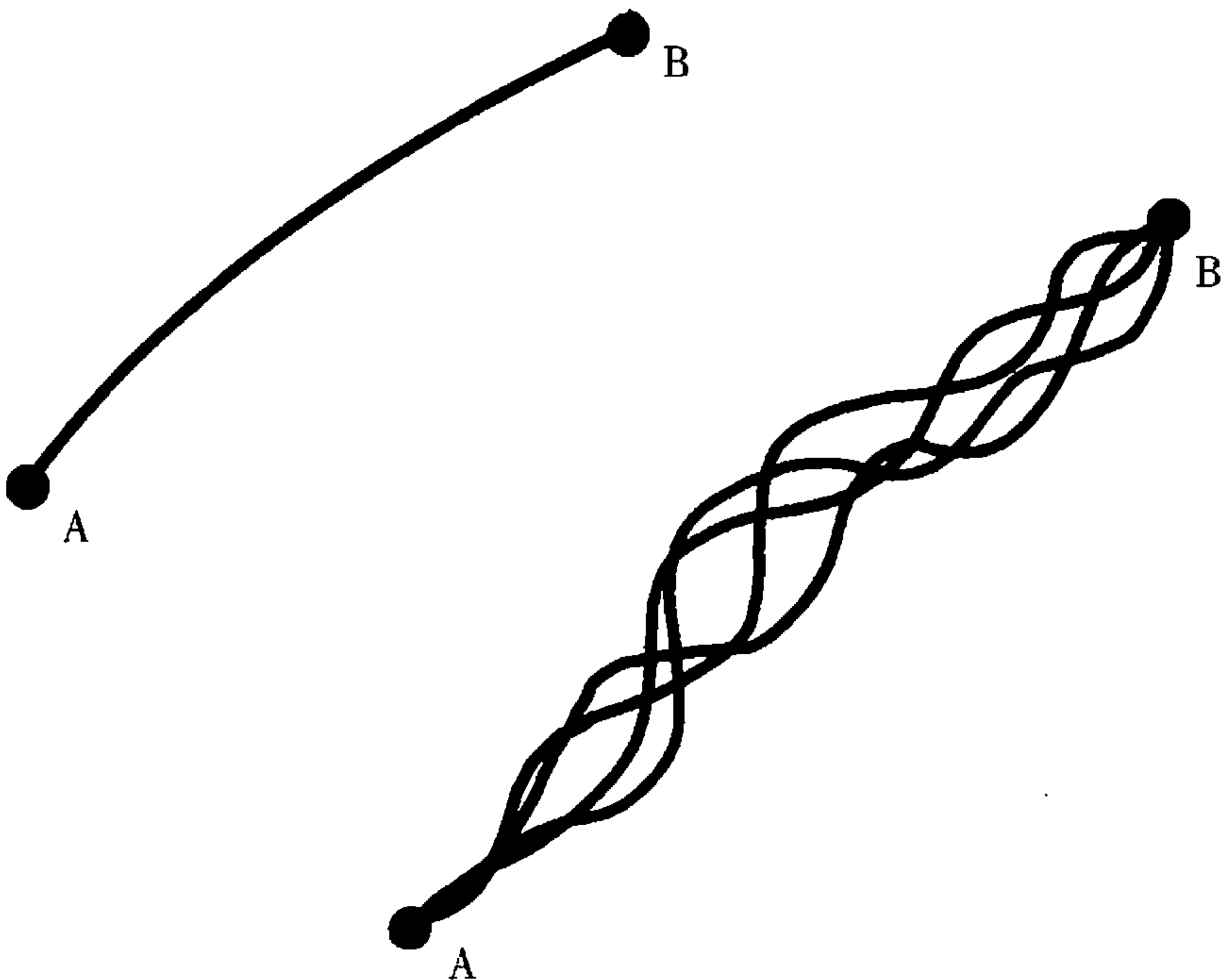
费恩曼(Richard Feynman)，加州理工学院一位得过诺贝尔奖的物理学家，以一种相对论专家用过的时空图把这种观点用于研究粒子

的历史。这种图中的一个轴表示时间，另一个轴表示空间。图中的曲线即所谓世界线表示粒子的轨迹或历史，有些世界线可以被排除，因为它们要求粒子运动得比光更快，但是许多世界线都表示粒子从A到B的正当路径。例如，对那个双孔实验，就可以想象画出一张图，其中有一个电子从发射枪起穿过某一个孔到探测屏幕上一个特定点的所有可能路径。有的路径是直截了当的，有的则是弯弯曲曲的。这种费恩曼图包括了时间和空间，图上有的路径表示电子快速通过实验装置，有的则表示慢速通过。每条路径，不论是直接的还是迂回的，快的还是慢的，都有一个可以计算出来的确定概率（严格地说是"概率幅"）与之相联。用以量度振幅的是一个称为作用量的量，它是能量乘以时间，正好是普朗克常量的单位。

世界线的概率幅并不全都"步调一致"，它们像池塘里波浪的振幅一样也能互相干涉，增加某一条路径的强度而抵消掉其他的振幅。这也就像波包中的波在除了一个小区域 Δx 外的所有地方都互相抵消。费恩曼的工作[4]表明，所有对应于可能的粒子路径的振幅这样干扰的结果，是去掉了其他的可能性，而只留下那些与经典力学中从A到B的轨道很接近的路径。当这种方法运用到双孔实验之类的实验时，所得到的结果与用薛定谔波方程所得到的完全一样。

实际上，这种方法的充分运用和得出详细结果还只限于一些很简单的特别情况。可以想象，对从电子枪到探测屏幕上每一个点的所有路径概率的计算是多么复杂。在大多数情况下路径的数量是如此之大，以至于严格地运用费恩曼的方法实际上做不到。但是它所蕴含的概念，以及它能作出与薛定谔形式的量子物理学相同的预言这一事实，都有着基础性的重要意义。[5] 费恩曼认为，在双孔实验里不仅要考虑一个电子同时穿过两个孔，而且要考虑它同时穿过两个孔的每一

条可能路径。通常的量子观点是没有轨道，而费恩曼的观点则是必须考虑每一条轨道。



插曲图 2.6 经典牛顿物理说，粒子沿着一条唯一确定的轨道从 A 到 B。费恩曼的量子力学解释却说，必须计算从 A 到 B 的所有可能途径的贡献并加在一起。这就是“路径积分”或“历史求和”方法。它能解释的事情之一就是一个电子怎么能“通过”杨氏双缝实验(插曲图 2.3)中屏板上的两个孔并自己干涉。

这种描述粒子轨道的方法称为“路径积分”，这是因为它把不同的可能路径相加，有时也更豪气地叫做“历史求和”。这后一个名称反映了对量子定则的一种解释背后的思想，它远离今天的主流，但是我很喜爱它并在《薛定谔之猫探秘》中作了介绍。它很符合由普林斯顿大学的埃弗里特(Hugh Everett)于 1957 年首创的所谓“多种世界”的说法。

埃弗里特注意到，对量子力学方程可以完全正当地这样解释，即宇宙每次都在量子层次上面临“选择”，它可以一分为二，而两种可能性都被选取了。在那个电子由 A 点经过有两个孔的屏板到 B 点的

实验里，按照量子理论，除非我们全程跟踪，否则就不可能知道电子穿过哪个孔，这也就是，说它走的是哪一条路是没有意义的。电子的“真实”轨道是由两条可能路径相加而给出的。但是经典理论说，电子有一条确定的路径，即使我们不去看，它也必定只穿过一个孔。当然，若我们想要看电子是穿过哪一个孔，那个特指的不确定性就消失了，我们做的是一个不同的实验，在这个实验里我们知道粒子是走哪一条路径。但是，埃弗里特说(或者说是方程式说)，对于每一个想看并看到了电子穿过某一个孔的观测者，都有一个在另一个世界里的观测者想看并看到电子是穿过了另一个孔。两者都同样真实。或者来看遇到偏振片的光子，每次光子都面临 50：50 的选择，宇宙分成了两个。在一个宇宙里，光子穿过了偏振片，在另一个宇宙里，它没有穿过。量子理论的这种解释的最奇特之处是，它对所有可进行的实验的可观测结果都能作出与概率解释完全一样的预言。这既是这种解释的优点(一个好理论毕竟得与所有已做过的实验相符)，也是它的弱点，因为大多数理论家都宽慰地说，既然它没有可检验的新预言来区分它自己与通常的解释，那就还是用概率吧。的确，概率使得电子可以继续被想象成一种点状粒子，如果你真想保持这种想象的话。

跳出油锅?

执教于泰恩的纽卡斯尔大学的英国物理学家戴维斯(Paul Davies)，在他写的一本量子物理学大学教科书里就极力主张这种观点。[6] 他在书中写道：“要不惜任何代价抵制诱惑，不要把电子想象成拉散展布在空中的波纹。电子本身并不是波，而只是其运动方式受波的原理支配。物理学家仍然把电子看作一种点状实体，只是那个点的精确位置不很确定。”他接着用社会上的犯罪浪潮作比拟，来描述那

决定着电子的可能去向的概率波。“犯罪浪潮不是那种有实物在振动的波，而是一种概率波……它有自己的动力学，像时装或是失业那样会起起落落，但一次犯罪当然仍是发生在一个地方，变动的只是那个抽象的概率。”

为着许多目的，尤其是大学物理教学的方便，物理学家的确把电子作为一种“真实”粒子来处理，而把与之相联的波看作“概率波”，这种波能相互干涉，能在通过小孔时衍射，还有其他所有的波能耍的花招。戴维斯在他的书里对学生们说：“是概率有着类似波的行为，而粒子本身则保持为小颗粒，尽管是难以捉摸地隐藏在指引其进程的波里……这种波粒二象性的哪一面显示出来，取决于所问的是何种问题。”这恐怕是很糟的讲授。若是麦克斯韦还活着，你可以试试问他，对把光波看成只是指引光子这种小颗粒运动的概率波有何感想，他的反应可能会很有趣。不管怎样强烈地试图坚持把电子或光子或其他什么想象成为粒子，这个概念总站不住脚。

以自旋为例。电子和其他亚原子粒子有一种属性，被物理学家称作自旋。除了别的作用之外，它对决定电子如何在原子中安排有着基本的重要性，而且量度它的单位也与量度陀螺自转或地球在空中自转时所用的单位一样。但是相似也仅止于此。电子的自旋只能朝着两个方向之一，即“上”或“下”，绝不能“侧向”或朝其他任何方向。自旋也像能量一样是量子化的。基本粒子的自旋可以用 h 为单位来量度，于是电子的自旋是1/2，即+1/2或-1/2，决不会有别的。所有那些我们喜欢看作“真实”粒子的基本粒子，如质子、中子、电子，都有半整数的自旋，即1/2，3/2，5/2等等。这些粒子都服从一种称为费米—狄拉克统计的统计规则。光子的自旋是1，而所有自旋为整数(1，2，3等等)的粒子都服从一种称为玻色—爱因斯坦统计的

不同规则。所以光子与电子是根本不同的。

像电子这样的粒子(统称为费米子)的最重要特征是不相容性。用原子中能级“阶梯”的概念，这是指只有两个电子能处在阶梯的同一格上，一个自旋向上，另一个自旋向下。[7] 同一格上不能再容纳第三个电子，因为所允许的两个状态已被占满。光子和其他整数自旋的粒子(统称为玻色子)就不那么势利，把它们随便塞进什么地方都行。另外，费米子是守恒的，宇宙中费米子的总数保持不变，而玻色子就比较短命。一开灯就能造出光子，光子又能被原子吸收并交出能量而消失。

所有这些都难以与由概率波指引的“真实”小颗粒的概念相调和。更难的是基本粒子的自旋还有其他一些怪异特征。例如，设想一个电子在旋转，那么它必须转过不是一个而是两个 360° 才能回到起始状态。虽然我刚刚讲过整个宇宙中的费米子是守恒的，但只要能得到足够的能量仍然可以制造同等数量的粒子和反粒子。例如，一对正负电子，在对宇宙中费米子计数时只能算是零。粒子与反粒子相互抵消。只要有能量就能制造正负电子对，就如在宇宙火球中制造它们一样。那么今天从哪里去得到能量呢？如果你很循规蹈矩，就会想到用大的加速器，如欧洲核子研究中心(CERN)的那一个，把粒子猛撞到一起而获得能量。但是你还可以更富于想象一点。按照不确定关系，可以在很短时间内“借取”能量来制造粒子，不过时间一到那些粒子又会再次消失。

仍考虑电子。一个电子的质量是 m，制造一对正负电子所需要的能量就是 $2mc^2$，即大约 1MeV。量子物理定律允许这样一对粒子在真空中突然出现并存在极其短暂的时间(即普朗克常量除以 1MeV)，然后互相湮灭而重归于消失。这样的粒子对叫做“虚”粒

子。每一对虽只能存在极短时间，但是真空里却闹哄哄地到处是这种粒子对，它们在不断地产生、消失，新对不断地取代老的。至少这是量子物理所说的真空的样子。而虚粒子的存在对粒子物理方程有着直接的影响。没有虚粒子，方程就不能正确预言带电粒子间的相互作用，而把虚粒子的影响包括进来后就能做到。

那么宇宙中的任何粒子究竟又有多么“真实”呢？戴维斯对他的研究伙伴讲的话，就与对大学生讲的有所不同了。他在一本为纪念物理学家德维特（Bryce DeWitt）（附带说一句，也是量子物理学的多种世界解释的倡导者之一）60岁生日而出版的书里有一篇文章，用了一个颇有挑衅性的标题“粒子并不存在”。[8] 他的论证要点也与许多理论家的观点一致，是从我们不能看见、触摸或是感觉到诸如电子这样的基本粒子这一事实开始。我们所能做的一切是进行实验并作出观测，然后以这些观测和我们的日常生活经验为基础来得出正在发生着什么的结论。我们很自然地总是试图把从日常世界得来的概念，如“波”和“粒子”，加到亚原子世界上；但实际上我们对亚原子世界所知道的一切，只是当我们以一定方式去碰它时就得到一定的反应。戴维斯在他文章的开头这样写道：“我确实试图推翻的正是所谓的朴素实在论。”他的结论是：“粒子概念纯粹是一种理想化模型，它在平直空间量子场论中有着某种效用。”他还说，粒子概念是“模糊不清的”，最终“必定被完全抛弃”。

问题在于我们还没有任何更好的东西去取而代之。但是在我们即将进入粒子世界时，需要记取的正是这来自现代研究前沿的告诫言词，而不是大学教材里那个概率波指引的真实小颗粒的安慰性说法。在过去50年里，物理学已经展示了那个居住着各种奇特客体的亚原子世界的胜境。我们把那些客体叫做粒子，是因为还没有一个更好的

名称。它们实际上是什么，我们并不知道。已有的最好理论用这些神奇客体即“粒子”之间的相互作用解释了过去的实验，并对新实验的结果作了预言。通过观测高能相互作用，物理学家得出描述这种作用的规则，然后又用来预言下一个实验的结果。好的理论能作出正确的预言，而最好的理论则能计算出宇宙怎样出现并爆炸成为现在的样子。但这并不意味着已经到达终极真理，也不意味着在原子内或在大爆炸发生时“真有”那种轻捷活跃的小而硬的粒子。所有那些理论里的真理只存在于数学公式中，粒子概念就像一支拐杖，我们这些凡夫俗子借它之助来走向对数学定律的理解。那些数学定律描述的，是复杂得难以置信的力场和弯曲到自己上面的时空，还有在你密切注视时会逐渐消失在虚粒子和量子不确定性的泡沫里的“真实性”。

粒子和波是我们现有的最好概念，表述现代物理学和对宇宙认识的巨大进展的最直接方式，仍是运用粒子概念。但这个概念仍只是对我们所不能恰当地领会或理解的事物的一种隐喻，我对不得不使用它而预致歉意。我觉得这就像一个盲人，试图给另一个盲人解释颜色的概念，凭借的却是一套建立在触觉基础上的理论。令人吃惊的倒不是我们的理论有缺陷，而是它竟然能够奏效。

注　释：

1. 电子是在1897年才由英国物理学家J·J·汤姆孙(J.J.Thomson)发现，所以普朗克在解释黑体辐射时，关于原子内带电粒子的性质和怎样“振荡”产生电磁波是有点模糊的。汤姆孙证明，电子是原子的组成部分，虽然不同的元素是由不同的原子组成，但所有原子中的电子都是一样的，都带有一个基本单位的负电荷，现在记作 e。他以这一成果获得1906年的诺贝尔奖。

2. 我们怎么知道电子波当我们不看它时在干什么呢？通过在多次实验中反复观察当我们看它们时电子是在哪里和沿什么路径运动，物理学家们认为波函数的正确数学描述是与这个很简单的说法相一致的。详情请见《薛定谔之猫探秘》。

3. 有趣的是，对很小世界的崭新认识源于20世纪20年代后半期的突破，而几乎是在

同一时间，对很大世界即宇宙的观测正在揭示其广延性，而且通过证实它的膨胀提出了它的起源问题。我不认为这里面有什么深刻意义。量子革命的到来可以比宇宙学革命早上 10 年或晚上 10 年。但它们都是其时代的产物，部分地依赖于制造大望远镜和用 X 射线探测原子结构的技术，而且它们也都大大得益于光谱学。所以两者都被制约着或多或少发生在同一个科学时代。

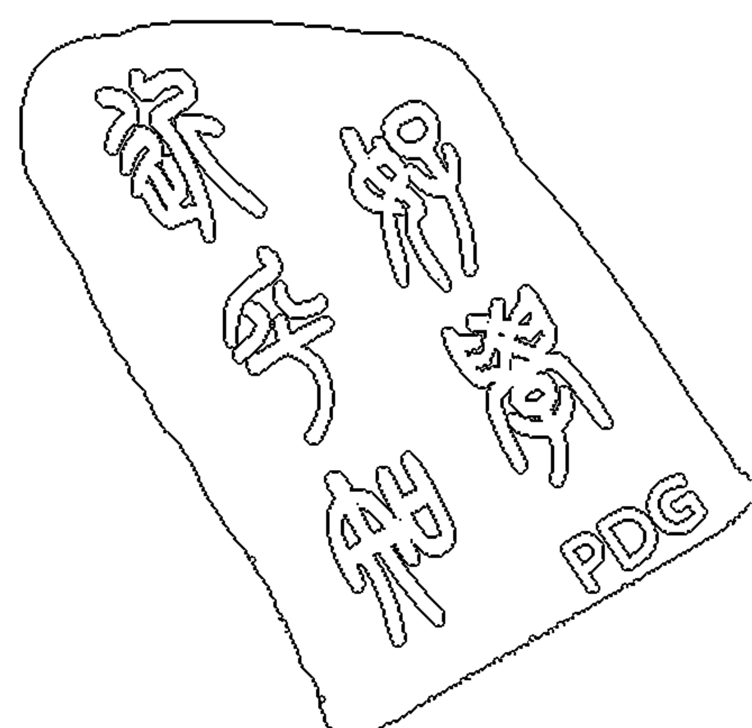
4. 其技术细节方面的详细描述见 Feynman & A. R. Hibbs，*Quantum Mechanics and Path Integrals* (New York： McGraw-Hill，1965)。

5. 可以把这种方法加以改进，使之有广泛的运用而不必详细计算每一条路径。例如，费恩曼已经证明，对应于经典轨道的最可能路径，是有着最小作用量的那一条，而且只有那些接近于最小作用量线的路径才需要包括到计算之中，因为其他路径的概率幅必定互相抵消。双孔实验的问题是，穿过不同孔的路径可以有相同的“最小作用量”，这样的路径在计算电子的命运时有着同等的重要性。

6. *Quantum Mechanics* (London： Routledge & Kegan Paul，1984).

7. 这里说的是过于简化了。按照我不拟在这里细说的道理，原子中的能级有这样的情况，即阶梯中的 4 格，或 4 个分立阶梯中各有一格，实际上是并列的，所以可以有 8 个电子处在很相似的状态。但是严格说来这些电子中仍然没有两个是在完全相同的状态。

8. *Quantum Theory of Gravity*(Bristol： Adam Higler，1984).



第三篇

大爆炸之前

如果我们不满足于实验事实的单调积累，如果我们要作出任何演绎或推广，如果我们追求的是有指导意义的理论，那么某种程度的推测总是不可避免的。

——爱丁顿

第七章 粒子和场

在20世纪30年代初，除了如何准确地解释量子理论中的概率和不确定性还是个谜外，看来物理学家对世界的组成已经有了很好的了解。有四种粒子，即电子、质子、中子和光子，还有两种知之已久的基本力，即引力和电磁力。质子和中子一起组成位于原子中心的核，电子则占有核周围的更广阔空间。由于质子带正电而电子带负电，因而原子是电中性的；由电子的不相容性造成的它们在不同能态上的量子物理安排，给了每种原子以独特的化学性质。

在物理学家按照不确定原理来察看原子的尺度时，量子理论的威力便清楚地显示出来。原子的直径约是一亿分之一(10^{-8})厘米，电子的质量是9×10^{-28}克。束缚在原子中的电子所具有的能量可以用玻尔的量子物理学早期形式来计算，也就是把电子看成在绕核轨道上的粒子，于是在数学上就与对围绕太阳轨道上行星能量的计算基本一样。这种朴素的计算给出，一个典型电子“在其轨道上”的速度约是每秒10^8厘米。上述数据合起来得出原子中典型电子动量的近似值是10^{-20}

克·厘米/秒，或者稍小一点。如果电子有比这大得多的动量(从而也有很大的能量)，它就会从原子中逃出，因为电子与核之间的电力将不足以束缚住这种高能电子。所以动量 p 的不确定值应该小于 10^{-20} 克·厘米/秒。再乘以电子位置的不确定值即一个稍小于 10^{-8} 厘米的量，结果就是 $\Delta p \Delta x$ 不大于 10^{-27} 的数倍，很接近于 $h/2\pi$ 的值。海森伯不确定关系决定了原子所能具有的最小尺度。如果原子比这个尺度还要小，那么电子位置的不确定值就会更小，其动量从而能量的不确定值就会更大，有些电子就会有足够的能量从原子里逃出来，就像 α 粒子由“隧道”逃出原子核一样。量子物理能够解释或者说能够预言原子的尺度，其所以能这样是“因为”在量子理论出现以前事情本来就是这样。

质子和中子的质量比电子的大得多，所以即使它们的速度较小，它们的动量($m \times v$)也有更大的不确定性。正因为动量较大，它们就能被限制在一个较小的范围(Δx)内，从而不确定值的乘积($\Delta p \Delta x$)仍然大于 h 。所以核比其周围的电子云要小得多，其准确尺度也再次能由量子不确定性来预言。

在 20 世纪 30 年代初，自然界看来就这么简单，物理学似乎已经找出了自然界的基本结构砖块。但是，只过了几年，世界就开始显得复杂得多；过了 20 年，物理学家就已证认出像化学元素那么多种类的“基本粒子”。经过一番革命性的新探讨，才给这些激增的粒子建立了秩序，并且用少数几种更基本的单元来解释它们的存在。对那些喜欢运用粒子概念的人来说，正如原子核被设想为由质子和中子组成一样，质子和中子本身以及其他粒子，现在都被认为由夸克组成。但在过去 50 年里粒子概念本身也经历了戏剧性的变化。正如光子被看作是电磁场的具体表象一样，电子和其他粒子也能被认为是它们各自

的场的具体表象。代替那多种场和粒子相互作用的思想，宇宙可以被认为是仅由各种相互作用的场组成，而粒子只是按照波粒二象性和不确定原理的规则出现的各种场的量子。所以，在讲述那些新进展和物理学家在过去50年里如何改变对世界的看法之前，考查一下物理学中场这个概念所表示的意思是恰当的。

场论

关于场是传递电力的媒介的基本思想可以追溯到1791年出生于英国纽因顿的先驱者法拉第(Michael Faraday)。法拉第有着非凡的经历，这里不妨稍稍离开主题，简略介绍一下他如何成了维多利亚时代的一位伟大的科学家和伟大的科学普及者。

法拉第的父亲是一名穷铁匠，他只受了那时穷人所能受到的基础教育，13岁就离开学校去跟一名书籍装订工当学徒。但法拉第至少学会了识字，而且有着很强的求知欲。他废寝忘食地阅读那些本是交给他装订的书，14岁时就因《大不列颠百科全书》上的一个条目而对电着了迷。他也读化学书，还在那很有限的资源允许的范围内自己做实验。1810年，法拉第加入了城市哲学学会，用业余时间去听物理和化学课，1812年21岁时，他的生活发生了变化，因为他去皇家学院听了由戴维(Humphry Davy)讲授的系列课程，戴维是一位大化学家，也是有电灯之前矿工用的安全灯的发明者。

法拉第深受戴维课程的吸引，他作了大量笔记，并自己装订起来永久保存。[1] 他渴望能用全部时间做科学研究，曾写信给皇家学会主席请求指点和帮助，但是没有收到答复。1812年他学徒期满，自己当了书籍装订工。他命运的改变是由于一次偶然事故，戴维因一次化学爆炸造成眼睛暂时失明，就叫那个正渴盼着的学生法拉第在自己的视

力恢复之前做助理。法拉第干得挺好，在戴维眼睛好了以后还把那装订好的听课笔记送了上去。戴维对此印象深刻，于是在几个月后即到了1813年，当皇家学院有一个助理空额时，就把这个位置给了法拉第。法拉第不失时机地接受了，即使那份工资比他当书籍装订工挣的钱还少。

法拉第此后的事业就全在皇家学院，1825年当了实验室主任，1833年成为化学教授。他是一位伟大的实验家和解释者，而不是数学家，他还作名副其实的科学普及演讲并且很成功，他在皇家学院首创的专以少年儿童为听众的圣诞节讲座一直延续至今。他于1867年去世时已经是皇家学会会员，并被公认为当时的科学巨匠之一。他又很谦逊，谢绝过爵士封号，还两次谢绝皇家学会主席之职。为找到一种描述电力和磁力作用的方式，他提出了一个概念，今天应该说是一个模型，就是“力线”，后来由麦克斯韦精心加工而用于第一个场论之中。

看看电荷之间的作用力，就知道这个概念很容易理解。同种电荷互相排斥(正电排斥正电，负电排斥负电)，而异种电荷互相吸引。法拉第的力线可以看成从宇宙中每个带电粒子延伸出去的数学上的线。每根线都起始于一个带一种电荷的粒子，终结于一个带相反电荷的粒子。像被拉伸的橡皮带一样，这些线试图把异种电荷拉到一起；但又像被压缩的弹簧一样，被挤压的力线往外反弹而使同种电荷分离。力线不仅能很形象地描述电荷间的作用，而且表现出一定的物理真实性，因为一个带少量正电的微小“检验”粒子放在两个带异种电荷的静止大物体之间时，将沿着一条最小阻抗的路线从带正电物体向带负电物体运动，这条路线(忽略其他力如引力)正是法拉第力线中的一条。

充满空间的所有力线就组成电场。在法拉第和麦克斯韦的经典场论里，场是连续的，就是说力线之间没有“空隙”，力线本身也没有间断。所以把场看成充满宇宙的弹性媒质更为恰当，电力和磁力就通过这种媒质传递。这就是“以太”的概念。这在维多利亚时代的科学家们看来是很自然的，因为他们熟悉的是机械物体的性质和实用工程上的成功。但是，随着相对论和量子论在20世纪的建立，这种机械的宇宙观被抛弃了。场就是场，不是什么弹性硬物，不可能用日常的思维方式来充分理解它。对场是什么的最广泛采用的“解释”，只有在你抛开成见去思考时才能有深切的认识。

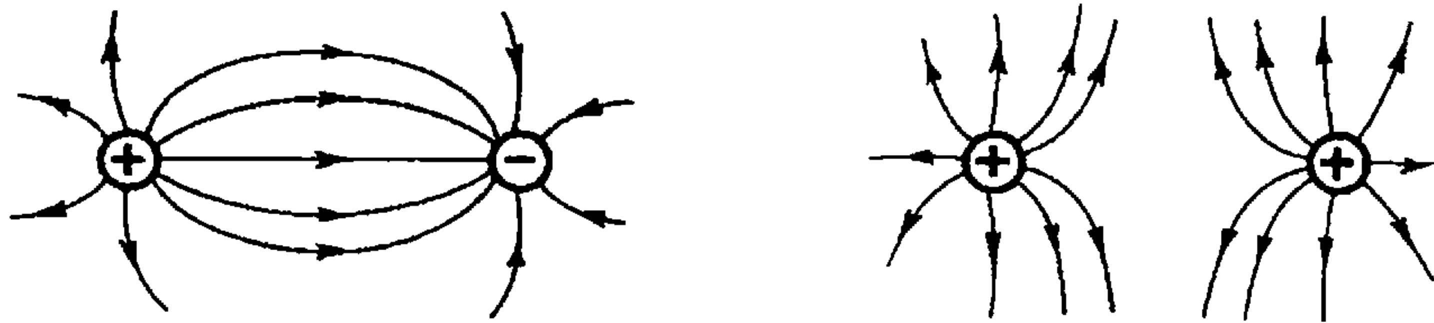


图7.1 由法拉第首创的力线概念能用来形象化地认识同种电荷相斥和异种电荷相吸。

在场论建立之前，诸如带电粒子或磁体之类物体间的相互作用，似乎是跨越它们之间的空隙来进行的，这就是所谓超距作用。场论则指出，作用是一种局域现象，每个带电粒子都与场相互作用，而场又与每个粒子相互作用，作用的方式则有赖于场与所有其他粒子间的作用。常用作比拟的是弹簧。如果拉一根弹簧的两端，它会伸长；如果挤压，它就会缩短。场就像这样。它能被延伸或压缩，它连接着两个粒子，就像弹簧中的物质把弹簧的两端连接在一起一样。这个比拟很好懂也合乎实际。但弹簧中的物质是什么呢？是原子的组合。而原子是怎样相互作用的呢？主要是通过电磁力。当弹簧被拉伸时，原子分离得更远；当弹簧被压缩时，原子被推得更靠近。按照这个比拟，电

磁场的拉伸或挤压就像是把原子拉开或推近，换句话说，就像电磁场拉伸或挤压自己！说到最后，或许最好还是用方程式来描述粒子与场的相互作用，而不要过分地借助比拟来想象。

现在的情况下这就是麦克斯韦方程组。麦克斯韦给出了第一个完备的场论，它既适用于电也适用于磁，并且指出磁就等价于运动着的电。这是一个经典的(意思是非量子的)电动力学理论。爱因斯坦广义相对论描述的是另一种日常生活中熟知的力，也就是引力，并且是另一个经典的(同样是在非量子的意义上的)场论。按照老的经典观念，世界只有这两种基本成分：实物和连接实物的场。

但是现在场已经成为物理学中终极的基本概念，因为量子物理学指出，粒子(即实物)本身只是场的表象。量子物理学使人们大吃一惊的首批事件之一，就是诸如电子这样的粒子必须得像波那样来对待。这是量子原理在这方面的首次运用，它告诉我们如何把物质波作为场来处理，每一种场对应着一种类型的粒子。例如，有一个充满宇宙的物质场，它由电子的波动方程来描述。[2] 但是，正如电磁波也必须被看成粒子这一发现所表明的，场能够直接造就粒子。的确，在量子世界里场必定会产生粒子。按照量子物理学，场中的能量不可能如经典图像中那样平滑地从一处连续地变动到另一处。能量是存在于称为量子的确定小包里，每种物质场都必定有自己的量子，每种量子都有一定的能量，或者说质量。粒子就是带有一些能量的场，由不确定原理限制在一定的范围内。光子是电磁场的量子；把量子原理以同样方式第二次运用于电子的物质场，我们会觉得是再次发现了电子可以被看作电子物质场的量子。把粒子解释为“场量子”，就称为二次量子化。它告诉我们，宇宙中除了量子场外一无所有。我们对量子场知道得越多，对宇宙也就认识得越好。

场有许多种。例如，有些场具有内在的方向，被称为矢量场，而其他场则没有。没有内在方向的场称为标量场，表示房间内各处温度的场就是一个例子。这个场显然充满了房间。在房间里任何地方放一个温度计，就能读出温度，靠近门口较低，因为那里有冷空气流进来，而靠近辐射源处就较高。但是温度计绝不会因与场的相互作用而被推着朝向或离开辐射源运动。这个场只有量值而没有方向。另一方面，电场是一个矢量场，可以在任何地方量度它的强度和它的方向。场中一个带正电的小物体会试图“沿一条力线”离开最邻近的正极并靠近负极。[3]

量子场还有一个重要的区分。迄今为止都是以或多或少相同的方式谈到电子和光子以及与它们相关的场，但实际上这二者之间有一个基本差别。电子属于费米子家族，其成员全都有半整数自旋，除了以正反物质对湮灭外，今日宇宙中的费米子既不能被毁灭也不能被创造出来。而光子却是玻色子，所有玻色子都有零自旋或整数自旋，可以被创造和毁灭，也没有不相容性。所以自然界有两类根本不同的场，即费米子场和玻色子场。我们过去惯常所想的粒子与力之间的区分看来就是由此而来。

按照老的图像，两个粒子相互作用是因为它们之间有一个力。这个力可以用一个场来表示，而这个场又可以转过来通过二次量子化以粒子来表示。按照新的图像，两个电子在相互靠近时相互推斥，是因为它们之间交换了一个或多个光子。光子是这个或那个电子周围电场的表象。它由不确定原理借获能量，突然出现，冲撞第二个电子，然后消失。第一个电子则在那个光子离开时受到反冲，结果就是两个电子相互推斥。[4] 一类场对应着费米子，产生出实物世界；另一类场则对应着玻色子，产生出相互作用，使实物世界能保持在一起，有时也

将那个世界的小部分碎裂开来。

电子周围的电磁场能产生虚光子，它们很短命，也不能离家很远。按照不确定原理约略估量，虚光子离开电子只能到它自己波长的一半处，然后就必须返回并被再吸收。波长较长则能量较低，所以较低能量的虚光子能离开电子较远。结果就是这样一幅量子图像，电子作为一个带电区域沉浸在虚光子海里，越靠近电子本身，海的能量也越高。

虚光子(实际上，还有普通光子)也能以虚正负电子对的形式产生电子，尽管它们也很短命，而且只能存在于不确定原理限定的范围内。那些虚电子也有自己的虚光子云，如此等等以至无穷。存在于电子周围，或是当一个电子被另一个推斥或散射时必定包含的这种活跃泡沫，与大多数人脑中同“真空”一词相连的平静景象大不相同。通过把量子理论的原理以这种方式运用于电磁场，物理学家已经得出一套量子电动力学理论和一个生气勃勃的真空，以描述电子、光子和电磁场在量子意义上的相互作用。这个理论被简称为QED，是现代科学的重大成就之一，它对解决电磁作用问题是如此成功，以致被看作所有量子场论的典范，并被用作构造解释其他相互作用的新理论的指南。但是，它有一个关键性的缺陷，这与那存在于每个电子周围的虚粒子云有关。

按照量子物理学，电子被虚光子的云所围绕，这些光子中的任一个或是全体，在被电子再吸收之前都可以转变成虚正负电子对或其他虚粒子对。所需的额外能量总可以按照不确定关系从场中借取，于是对这种产生虚光子和虚正负电子对的复杂连环简直就没有任何限制。当量子物理规则被一丝不苟地运用于计算这些虚粒子的连环能包含多少能量时，结果竟没有什么限制，电子加上其周围虚粒子云的能量成

为无限大。由于电子和它的云不可分割，所以乍看之下这似乎意味着电子必定具有无限大的质量。

绕开这个困难的办法看似胡闹，实则有效。从数学上讲，电子周围云的无限大质量可以这样来补救，即假定“裸”电子(如果能有这种东西)有着无限大的负质量。小心地玩这种数学把戏，就能让两个无限大相抵消，而留下一个质量就是已经测出的电子质量。这一招叫做重正化。它有两点令人不满意。第一，它实际上是在一个数学方程式的两边同除以无限大，我们大家全都在学校里听老师讲过这是不允许的。第二，即使如此它仍不能给出“正确”的电子质量。重正化能给出一个有限的质量，但可以是任何有限质量，物理学家不得不选择一个正确的并亲手将其塞进去。他们能够解出方程，只是因为他们已经知道了在寻找的答案。但是奇迹出现了，在人为地塞进了仅此一个关键数值后，方程式就能很精密很准确地“预言”其他许多至关重要的参数值，这就是这么多物理学家很高兴与重正化一起过日子的缘故。

包括了重正化的 QED，的确能在量子物理学的框架内很好地解释带电粒子和电磁场的行为，竟使得大多数物理学家觉得不必对那些问题想得太多。但是如果他们能得出一个理论，在其中无限大能够自行抵消而无需重正化，他们的欢乐才将真是无限的。量子电动力学是已有的最好和最完整的量子场论，但是仍不完美。对一个将能解释宇宙中的所有相互作用，乃至解释宇宙本身的存在的完美理论的追求，现在正开始结出果实。但在真能品尝那首批果实之前，还得先跟上时代的进展，去看看那座名副其实的博览园里的展品，那里有过去 50 年里发现的新“粒子”，还有为完成对量子场种数的统计所需要的两种新“力”。

两种新力和一个粒子“动物园”

20 世纪初，科学家知道 92 种不同化学元素的存在。每种元素都由一种特定的原子组成，于是这 92 种不同的原子就被认为是自然界的基本构造单元，尽管自然界需要这么多的“基本”构造单元似乎是太浪费了。由于门捷列夫(Siberian Dmitri Mendeleev,1834—1907)的开创性工作，19 世纪后半叶的化学家开始认识了不同重量的原子之间的关系。门捷列夫发现，当把元素由氢开始按原子重量递增的顺序排列时，每经过某种有规则的间隔就会有化学性质相似的元素出现，结果就得到一张周期表。在这张把元素按化学性质归类的表上有着一些空格，尽管门捷列夫并不知道这种周期性格局的成因，但他单凭周期性的简单事实就能预言那些空格将由新发现的元素去填充，并且预言了那些新元素的性质和原子重量。后来发现的“新”元素严格证实了他的预言。

到物理学家能够把原子分开并揭示其内部运转方式时，他们知道了原子包含有三种类型的粒子，即电子、质子和中子。玻尔手中的量子物理学已经能够解释化学元素的观测性质，以及作为门捷列夫周期表根由的原子内部结构。周期表首次出现在门捷列夫于 1869 年发表的一篇论文里，玻尔的量子物理解释则在 20 世纪 20 年代，相差将近 60 年。但就在原子的本质通过亚原子粒子来得到认识时，亚原子粒子本身又开始显得难以捉摸了。

量子形式的电磁力完全足以解释带负电的电子的行为，但是在电中性的原子里电子只是核内带正电的质子的伙伴。那些都带正电的质子怎么能在核里聚在一起而不被电斥力推开呢？1871 年出生的新西兰物理学家卢瑟福在 1910 年确认，原子中的所有正电荷都集中在一个很小的核里。他和同事们在英国曼彻斯特大学的这一发现，是通过

观测轰击金箔中原子的 α 粒子如何偏转实现的。卢瑟福还在 20 世纪 20 年代初猜测到，质子必定还有一种中性对应物，他称之为中子，其质量与质子的大致一样，但不带电荷。

要解释为什么有些原子有着很相似的化学性质却又有不同的重量，就得有中子。化学性质是由原子中电子的数目决定的，这个数目总是与核里的质子数相同。所以要改变原子的重量而不改变其化学性质，就只有往核里增加或从中取出电中性的粒子，即中子。具有相同化学性质和不同重量的原子称为该元素的同位素。1891 年出生的英国物理学家查德威克，在同卢瑟福(那时已在剑桥的卡文迪什实验室)一起工作时，于 1932 年用一系列实验证实了中子的存在，因而获得 1935 年的诺贝尔奖。

从查德威克发现中子到他获得诺贝尔奖之间这短暂的三年，是亚原子物理显得很简单的时期，需要物理学家操心的似乎只是区区四种基本粒子。中子的存在甚至还有助于解释核的稳定性，因为带正电的质子可以在一定程度上由中子来隔开。但是这种帮助是不够的，随着人们认识到这一点，粒子世界的简单性就开始丧失了。

对简单性信念的第一击来自一位日本研究者汤川秀树(Hideki Yukawa)。汤川生于 1907 年(1981 年刚去世)，在京都大学和大阪大学上学后，1935 年起在大阪大学攻读博士学位，1938 年获得学位后又在该大学任教，1939 年返回京都大学当物理学教授。像其他物理学家一样，汤川也为原子核如何保持稳定所困扰。他推测，必定有一种比电磁力更强的力，即使在电斥力“要”使质子分开时也能把它们抓牢。但是在日常世界里看不到关于这种强力的任何证据，所以那必定是一种日常体验不到的新力，它只在很小的范围内起作用，把质子和中子保持在核内，而允许单个粒子(或是前面已讲过的 α 粒子)一旦

越出其作用范围就自由飞行。汤川用与电磁力的类比来描述这种新力。

按照电磁场论，力由于交换粒子即虚光子而产生。由于光子的质量为零，有着很长波长的光子所携带的能量就小得可以忽略。所以，原则上电磁力的作用范围没有限度，与一个电子相联系的虚光子可以作用于宇宙中任何地方的另一个电子，尽管那光子只有很低的能量因而作用很微弱，当然在两个电子靠得很近时其间的作用就强得多。

但若光子有质量会怎样呢？那就会有一个造出虚光子所需要的最小能量值 ΔE 。该能量包的有限尺度就会对虚光子的寿命给出一个符合海森伯不确定关系的限度 Δt 。由于没有任何东西能跑得比光更快，有限的寿命就意味着，这种可称为“有质量光子”的粒子只有一个很有限的活动范围，因为它在允许的生存时间结束之前必须要么回到其产生处，要么找到另一个粒子以埋葬自己。汤川于是在 1935 年提出，必定有另外一种场，与电磁场相似，但却与质子和中子相联系。这种场的量子像光子一样也是玻色子，但与光子不同的是有质量。这种玻色子只能由“感觉”得到强场的粒子来交换。汤川所说的实际上是，电子恰恰与这种强力无关。

汤川假说的美妙之处在于，能够计算出这个新品种的玻色子应该有多大质量。它的活动范围必定不会比原子核的尺度大很多，否则即使有不确定性的帮助 α 粒子也不可能逃脱，还会出现一些与实际观测不符的现象。对卢瑟福小组的开创性散射实验后来所作的细致改进表明，核的典型尺度只有 10^{-12} 厘米。汤川由这个单一测量结果和不确定关系算出，传递强力的粒子质量约为 140MeV，比电子质量大 200 多倍，但仍只有质子质量的 1/7。

怎么检验汤川的假说呢？那时物理学家还没有任何办法看到核的

内部以寻找新的玻色子。但既然强力取决于虚玻色子的交换，对应的真实粒子就应该能按照有关方程式在有足够能量的地方产生出来。[5] 例如，在宇宙诞生时的原始火球里，这种粒子必定就很丰富(但别忘了，在 1935 年还没有人真正想过什么宇宙的诞生；伽莫夫当时仍只是一位量子物理学家，他对大爆炸的热情倡导还是 10 年后的事情)。按照有关方程式，这种粒子应该是不稳定的，它们所具有的质量能量能够转换成别的稳定的形式。反过来它们也能由快速运动(即高能)粒子之间碰撞提供的能量制造出来。今天，物理学家用巨大的粒子加速器来使电子和质子束迎头对撞，或是撞到静止的靶子上，从而产生出阵雨般的短寿命粒子。这些粒子按照关系式 $E=mc^2$ 或写成 $m=E/c^2$，由碰撞粒子的动能造就。这是应该认识到的重要一点。“新”粒子并不是撞到一起的粒子破裂的碎片，而真正是新的，由纯粹的能量新造出来的。所以碰撞能够很容易地产生出静质量比参与碰撞的粒子更大的新粒子，只要碰撞粒子的运动能量大于所要求的静质量。[6]

在 20 世纪 30 年代，必需的高能粒子的唯一来源是宇宙自身，那就是一些轰击地球大气的高能质子、电子，还有(我们现在知道的)其他粒子，统称为宇宙线。一个宇宙线粒子与地球高层大气的另一个粒子碰撞，能够产生出新粒子，包括强力的玻色子。首批高能物理学家是这样一些观测者，他们找到了监测从实验装置中通过的宇宙线的办法，即用一个很像改进的盖革计数器的装置，让宇宙线在其中激起闪光，并在照相底片上记录下来，此外还有别的监测方法。一旦抓到了一个或一把飞驰的宇宙线粒子并拍摄下它的足迹，就可以根据其路径在磁场中的偏转方式来算出它的电荷，甚至可以由它的路径被磁场弯曲的程度来推算其动量(从而最终得出其质量)。

1936 年，高能物理学的先驱者之一美国人安德森(Carl

Anderson)在研究地面探测器中的宇宙线轨迹时，发现了一个比电子重但比质子轻的粒子径迹。看来像是找到了汤川的强力传递者，那种粒子被起名为 μ 介子，或简称 μ 子。但后继的研究很快表明 μ 子并不是强力的传递者，因为它的质量不对，也并不显示足够的与原子核发生作用的趋向。然而在 1947 年，另一位宇宙线物理学家英国人鲍威尔(Cecil Powell)发现了一种短寿命的玻色子，其性质正好符合要求，尤其是质量很接近于汤川预言的值，而且显得很急于与核粒子发生反应。它被称为 π 介子，或 π 子。而 μ 子原来是 π 子衰变的产物之一。汤川获得了 1949 年的诺贝尔物理学奖，他是第一个获此殊荣的日本人，而鲍威尔则于 1950 年被授予该奖。至于安德森，他已经在 1936 年也就是他发现 μ 子的那一年得了诺贝尔物理学奖，但不是由于 μ 子而是由于一个大不相同的发现，那发现初看起来只是为粒子“动物园”增添了一名成员，实则是使粒子种类一夜之间翻了一番。

狄拉克(Paul Dirac)，一位出生于 1902 年的英国物理学家，是 20 世纪 20 年代开始的量子革命的关键人物之一。他把由海森伯建立的第一个量子力学与爱因斯坦的狭义相对论融合起来，并在这样做的过程中引进了电子自旋量子化的概念(这个概念立即被用于其他粒子)；他发展出了一套非常完整的量子理论的数学表述，并就此主题写了一本有广泛影响的教科书，该书今天仍为学生和研究者们使用；他对 QED 的发展也作了重大贡献，尽管他直到去世(在 1984 年)都一直对重正化这件事深感不满，他觉得那就像是把一张纸糊在墙上遮盖裂缝。但对物理学家圈子以外的人们来说，狄拉克对认识宇宙所作的最著名的贡献是他 1928 年的预言，即物质世界的粒子都有以反物质形式存在的配对物，也就是镜像粒子。

奇怪的是，作为一个硕果累累和刻意求精的理论家，狄拉克关于反物质的预言几乎是偶然地作出的，而且首次公之于世时是以一种很不明确的方式。他发现自己正在研究的描述电子行为的方程实际上有两组解，而不是一组。任何人若是知道简单的二次方程，也就是含有未知量平方的方程，马上就能明白这是怎么回事。平方总是正的。2×2得4，-2×-2也得4。所以对“4的平方根是什么？”这个问题的“答案”是2或-2，两个答案都正确。狄拉克所对付的方程比这要稍微复杂一些，但道理是一样的。那些方程有两组解，一组对应着电子(带负电荷)，另一组则对应着一种未知的带正电荷的粒子。

1928年时，物理学家只知道两种粒子，即电子和质子，对中子的存在还只是一种强烈的猜测。所以狄拉克开始时以为他的方程的正电解必定是表示质子。看来即使是狄拉克这样的大物理学家，在20世纪30年代时也是在黑暗中摸索，他也看不出对应着正电和负电解的粒子必然具有同样的质量。只是到了现在，我们才能以事后的高明说，电子的配对物“当然”要有与电子一样的质量，质子质量太大，当不了电子的伴侣。起初也没有什么人认真看待狄拉克的对称解，更不要说协力去寻找猜想的新粒子了，即使在今天有一个类似的假说提出来，恐怕也会如此。当时的物理学家都不认为狄拉克的计算是在对世界作出有意义的揭示。方程存在的另一个解被置之不理，正像一名工程师在解二次方程时会去掉一个解，而只保留显然能用于手中工作的那个解，那工作或许是建造桥梁，或许是别的什么。

但到了1932年，安德森正在研究宇宙线，用的是一个云室，那个装置能使宇宙线粒子在其中留下轨迹，就像飞机在天空中拖出的那条凝聚物尾巴。那些轨迹可以被拍摄下来，对照片上的图样就能作从容不迫的分析。安德森做的事之一是研究宇宙线粒子的轨迹如何因磁

场的影响而变化，他发现有些轨迹的弯曲程度与电子的轨迹完全一样，但弯曲方向却正好相反。[7] 这只能表示那相应的粒子的质量与电子一样而电荷却相反（带正电）。这种新粒子很快就被证实正是由狄拉克方程所预言的粒子，它被叫做反电子，或更普遍地是叫做正电子。这项工作使安德森获得了诺贝尔奖。狄拉克则已于 1933 年与薛定谔分享了该项奖。

正电子和中子是在同一年被发现的（实际上在宇宙线轨迹中以前已经有过正电子存在的证据，但都被误认为是反向运动的电子的轨迹）。将狄拉克的计算扩展到原子中的所有粒子，物理学家就有 6 种粒子（另加上光子）要操心了，那就是电子和正电子，质子和（推测的）带负电的反质子，还有中子和（推测的）反中子。[8] 按照有关物理定律，一个粒子与相配对的反粒子相遇时，二者会湮灭而成为高能光子（γ 射线）的爆发。物质世界里的电子和正电子就可以这样相互抵消。方程式同样也可以反向进行，若有足够的能量可以得到，则正负电子对或别的正反粒子对也能被造出来。但在所有这些反应中每个粒子必定总是与它的精确镜像配对，而不可能是，比如说，质子配反中子。所有这些预测都已由实验证实，尽管反质子和反中子直到 20 世纪 50 年代中期才被探测到。所有这些物质与能量之间的转变总是服从 $E=mc^2$ 和量子物理定则，并且是大爆炸标准模型的基础，这在前面已经讲过。

正电子和中子是在同一年即 1932 年发现的。μ 子的发现是在 1936 年，π 子是在 1947 年。于是已经清楚，有两类粒子，一类是能感受强力的（质子，中子，还有传递强力的 π 子），另一类则不能（电子，其实 μ 子也是）。[9] 这是一种对所有粒子的新分类法，“物质”粒子和传递力的粒子都包括在内。能感受强力的称为强子，不能感受强

力的则称为轻子。所有的轻子都是费米子即有着半整数自旋。我们迄今讲到的轻子只有电子和 μ 子，μ 子除质量大得多外其他方面都与电子一样。强子中也是费米子(即“物质”)的又称为重子。质子和中子都是重子。在粒子之间传递力的玻色子已经明确是介子。π 子是一种介子，它又分为三种。一种是中性的 π 子，不带电荷。当一个质子与一个中子交换一个中性 π 子时，它们就聚在一起并保持不变。质子之间可以相互交换中性 π 子，中子之间也如此。还有两种分别带正电和负电的 π 子，它们互为反粒子。[10]如果一个质子把一个带正电的 π 子给了一个中子，那么质子就会变成中子，而中子就会变成质子。这与中子把一个带负电的 π 子给了质子时所发生的变化完全一样。这些交换和转变也都有助于质子和中子维持在一起。

仅是描述原子所需的粒子种数就已经大增了。但还有一种粒子和一种场要加到名单里来。

再回到 19 世纪末，卢瑟福先在剑桥后在加拿大工作时，发现铀会发出两种辐射，并研究了这些辐射的性质。[11]这些“射线”之一即 α 辐射，后来查明是氦核，即两个质子和两个中子束缚在一起的稳定态。另一种他称为 β 辐射，后来证实是电子。所以原子也能发射出电子。这些电子并不是来自原子核周围的云。卢瑟福和同事索迪(Frederick Soddy)在 20 世纪初证明，一个放射性原子发射了一个电子后会变成一个不同元素的原子。后来的研究则表明，在电子射出时核里的一个中子变成了质子，于是就成了一个新核，对应着一个不同元素的原子。实际上这个过程只对很少几种不稳定的核才会发生。大多数原子里的大多数中子都挺舒服地待在那里。但是一个离开了原子核的孤立中子会在几分钟内衰变成一个电子和一个质子。这个过程叫做 β 衰变，它还必须包括所有迄今已讲到的之外的另一种力和另一

种粒子。

历史上，新增的粒子总是由一些或一个物理学家先提出来。β衰变是20世纪头10年里物理学家的一个主要研究课题，他们的惊人发现之一是，这种衰变所产生的电子可以带着不同的能量出现。一个中子衰变所产生的质子和电子的总质量，比中子质量要小约1.5个电子质量。[12]这么多能量应该作为动能由质子和电子来分享。那个质子既是留在原子核内，当然就不会动得很多，那么几乎所有的额外能量看来都该归于电子，使它除静质量外还有动能，于是每个由放射性原子这样产生出来的电子就应该以很大的而又可预知的动能飞出。但实验结果却是，β衰变电子的实际能量总是小于应可得到的能量，有时还小得很多。那相差的能量到哪里去了呢?

泡利(Wolfgang Pauli),一位1900年出生于奥地利的瑞士物理学家，于1930年给出了一个答案。必定还有另一种粒子也与电子和质子同时产生出来了，这种额外的粒子带着“丢失的”能量没被看见就跑掉了。被交给此项任务的粒子必定质量为零，也不带电荷，不然的话就已经在实验中被注意到了。

这样一种古怪的猜测不会立即为物理学家欢迎。它看来太容易了，而且这种每当物理实验中有疑难现象就乞灵于一种新的难以探测的粒子来作解释的做法也确有风险。但是泡利固执己见，终于在1933年从一位比他小一岁的意大利物理学家费米(Enrico Fermi)那里得到了支持。费米接过了泡利的主意并给了它一个像样的立足点，他的做法是在计算中引入一种新的力，即所谓“弱”核力。

无论如何，场论要求有一种新力来说明β衰变。不可能是由强力来负责任(电子不“理会”强力),也肯定不会是电磁力或引力。费米让他的理论尽量向QED看齐，从而得出这样的认识，中子变成质

子时会发射一个粒子，那粒子是新场的载体，是带电的玻色子，记作 W^-。这种玻色子(现在叫做中间矢量玻色子)带走了电荷和额外的能量，而中子变成质子并后退。这种玻色子的质量非常大(早期不完整的弱力理论却不能得出那质量究竟是多大)。[13]它不仅带有造出电子所需要的能量，还有着从真空借来的巨大的虚能量，所以它很不稳定，不可能存在很长时间。它甚至不能活到与任何别的粒子发生作用，而是很快把借来的能量交还给真空，剩余的能量则形成电子和一种新粒子，就像高能光子转变成电子和正电子一样。电子是一种轻子，所以严格说来，要使宇宙中轻子的总数守恒，那个额外产生出来的粒子就必须是反轻子。(由于开始时是一个重子即中子，结束时也是一个重子即质子，所以重子数也是守恒的。)费米把那个额外粒子称作中微子，意思是“微小的中性粒子”，今天知道应该叫它电子反中微子。

1933 年，英国的《自然》杂志拒绝了费米陈述这些思想的文章，说它“过于推测性”。但他的文章很快用意大利文，不久后又用英文发表了。核反应堆会产生中微子的洪流，可以在一些实验中利用它们，这种粒子存在的证据就是这样在 1953 年得到的。它具有理论预期的所有特性(或者说缺乏特性)，尽管今天有人猜测，中微子可能具有很小的质量，比电子的质量还要小得多。

于是在 20 世纪 50 年代初，物理学家就有了足够的粒子和场来解释原子的行为。弱场和它所传递的相互作用，对核聚变和裂变过程，对恒星中元素的生成，对太阳高温状态的维持，直到对原子弹的威力，都起着至关重要的作用。电磁力保持其已为人熟知的地位，引力则一如既往地顽固拒绝到量子群落入伙。并不是所有的场都屈从于重正化的把戏，无限大问题虽在 QED 中被制服[14]，但在引力的情况解决同样问题的努力均遭失败。已知的轻子只有两种(还有它们的配对反

粒子)，即电子和 μ 子，每一种都联系着特定类型的中微子。于是注意力就集中到了由强力支配的粒子上。但在此后的 10 年里，物理学家对强子的性质探查得越多，他们所得到的图像就越混乱。

1932 年时，物质世界看似可以用 3 种粒子来解释。1947 年时，粒子的种类有了半打(另加上反粒子)。到了 1951 年底，至少已有 15 种“基本”粒子，而名单的扩大才刚刚开始。今天，粒子的种类比周期表上的化学元素还要多。在 20 世纪 50 年代那 10 年，物理学家每次开动新的粒子加速器都能造出新品种的强子，粒子“动物园”的成员越来越多，不是来自那稀薄的空气，而是来自单纯的能量。能量则来自机器，越来越大的机器，其中带电粒子如电子和质子被电磁力加速，然后相互对撞，或是撞到固体物质靶上也就是撞进原子核里，因为加速的粒子很容易穿过核外的电子云，就如炮弹穿过海上的薄雾。没有任何物体能被加速到光速，所以当越来越多的能量加进去时，粒子并不是跑得越来越快。一旦粒子的速度已增到光速的一个可观部分，它们就更多地是增加质量。而当它们碰撞即发生相互作用时，所有这些增加的质量都可以用来制造别的粒子(几乎都是短命的)，新粒子会在气泡室或其他探测器里留下足迹。一般说来，这样得到的每个新粒子都有一个反粒子伙伴相随，轻子数和重子数都是守恒的，而介子则可以随意制造。

我必须再次强调，绝不能把这些“新”粒子设想为原就存在于质子或其他任何用于碰撞的粒子“内部”。新粒子是由加进机器的能量造出来的。它们都已被命名，种族特征也已被证认和标记，有的特征的名称本身就很怪，如“奇异性”。粒子物理学的状况，很像门捷列夫之前的化学，那时元素都被证认了，其性质被确定了并互相比较了，却不知道那些性质和同类关系是怎么产生的。化学的新阶段是随

着元素周期表，以及后来依据原子内部结构对周期表的解释而来到的。粒子物理学新阶段的到来，也是由于20世纪60年代初粒子“周期表”的提出，以及几年后由强子的内部结构来对这个新周期表所作的解释。

八正法：混乱中的秩序

20世纪50年代末的场论在对强子多样性的认识上没有什么进展。有无限大的问题，与QED不得不用重正化来处理的类似，还有对每种粒子都需要单独引入一种场的困难，这在只有两三种基本粒子时还可以接受，但随着粒子种数增加到几打甚至100以上，就越来越使人焦虑不安了。大多数理论家在60年代初放弃了场论，试图找出强作用问题的另外出路。我不拟在这里介绍那些尝试，因为正是场论在70年代获得了成功。虽然寻找强子性质规律的动力部分地是来自50年代场论中的思想，粒子“周期表”作为门捷列夫表那样的分类系统，在粒子物理学的新阶段开始时却有其独特的价值。

这个分类系统是由两位物理学家，1929年出生的美国人盖尔曼（Murray Gell-Mann）和1925年出生的以色列人内埃曼（Yuval Ne'eman），各自独立地得到的。内埃曼的教育和专业发展曾被紧随第二次世界大战而来的中东战争打断，以色列正是在战火中作为一个国家出现在曾是巴勒斯坦的地区。在这个动荡时期之后内埃曼仍留在以色列空军部队里，在履行军事职责的同时还能找到机会做学问。虽然学士学位是在工程学上得的，他的兴趣却在20世纪50年代转到了基础物理问题上，那时他是以色列驻伦敦大使馆的武官，同时又在攻读博士学位，学位是在1962年由伦敦大学授予的。盖尔曼走的则是常规道路，从耶鲁大学到麻省理工学院，1951年在那里得了博士学位，

然后去了普林斯顿，接着是芝加哥大学(在那里曾一度与费米一起工作)，又于1955年到了加州理工学院。是盖尔曼把“奇异性”作为粒子的一种量子化属性引入粒子物理的，目的是为了说明20世纪50年代初观测到的一些高能相互作用的新现象。

奇异性是基本粒子看来应有的一种属性(或者说是我们把世界设想为是由粒子组成时需要放到模型里去的一种属性)。它的神秘性比起电荷来并不更多，也并不更少。有的粒子带电荷，有的不带，电荷分为两种，我们用+和-来标记。如果要更精确并把零电荷包括进去，就有三种选择：+1，0，-1。不同的粒子有不同的奇异性，可能的情况比电荷要多，但道理是一样的。奇异性可以是0，-1，+1，±2或更大。在粒子的强相互作用中奇异数必须守恒。正如中子转变成质子时必须产生一个电子来使电荷平衡一样(还有一个反中微子来保持轻子数不变)，在强作用中也必须有带适量奇异“荷”的粒子生成，以保持奇异数不变。这就使一些原来以为可以进行的反应被限制住了，物理学家在20世纪50年代正是得到了这个“奇异”的结果，从而也有了“奇异性”这个名称。

盖尔曼和内埃曼于是各自将新粒子和熟悉的老粒子放在一起，按照它们的电荷、自旋、奇异性和其他性质来分类。由于最先分出的一些类中包括8种粒子，盖尔曼就把他的分类法称作“八正法”，以赞颂佛教学说中的“八德”。后来才知道，归为一类的粒子可以是1种、8种、10种或27种，同一类中的每名成员都是某种共性基础上的变异。这套分类系统是1961年提出的，1964年盖尔曼和内埃曼两人合编了一本书，书名就叫《八正法》[15]，其中收录了他们自己的原始论文和其他有关粒子“动物园”新认识的重要著述。到该书出版时，这种分类法已经成功地预言过新粒子的存在，从而获得了正如量

子物理学出现之前的门捷列夫周期表那样的牢靠地位。

八正法扩展到把重子归入应有 10 名成员的一类时，该类却有一个空缺，需要增加一种粒子以构成完整图像。盖尔曼把那种粒子叫做 Ω^-，用的是希腊字母表里的最后一个字母。那个空缺“属于”一种带负电、奇异数为 -3、质量为 1680MeV 的粒子。纽约的布鲁克黑文实验室和日内瓦欧洲核子研究中心的研究者，果真于 1963 年找到了这种与盖尔曼的预言精确相符的粒子。门捷列夫周期表提出之后 60 年，才得以用完整的原子结构理论来作出解释。八正法得以用完整的强子内部结构理论来解释则花了 10 年多一点。本来可以不用这么久，但是许多物理学家开始时不愿意接受早在 1964 年就由盖尔曼和茨威格(George Zweig)各自独立地提出的思想，即“基本”粒子如质子和电子实际上都由叫做夸克的特殊粒子组成，夸克有三种，所带的电荷是电子电荷的一个分数，最后这一点听起来简直是异想天开。

夸克

回顾一下 20 多年前物质夸克模型的产生经过，恐怕连提出者开始时都没有很认真地看待它。认为质子、中子和其他粒子实际上都由另外三种粒子组成，三种粒子中有的所带电荷为电子电荷的 1/3，有的为 2/3，这实在是与自 19 世纪末以来所熟知的一切格格不入，以至于开始时只能作为一种手段、一种数学技巧提出来，以简化计算并为八正法的结果给出一个基础框架。这过去不是、现在也不是什么坏事。它再次提醒我们，所有关于基本粒子及其相互作用的模型都只不过是人为的辅助物，我们借以对所发生的事情获得一个图像，这个图像我们凭日常经验看来是熟悉的，或至少是可认识的。奇怪的倒是，随着夸克模型越来越好地建立起来，近年的许多关于粒子物理的论述

似乎忘记了即使是最好的模型也只不过是我们用以想象的手段，而开始提出质子、中子等等都是由“真实”的小颗粒即夸克组成，那些小家伙在我们曾认为是“基本”的粒子里面闹得正欢。这恐怕是重蹈了那个不准确的早期原子模型的覆辙，在那里原子也被认为是由三种小颗粒即电子、质子和中子组成。

不管其“真实性”如何，夸克模型的确很简洁地解释了粒子世界的相互作用。[16]质子、中子和传递强力的 π 子，全都可以由两种夸克来表述，夸克的名称可以任意选取，以作区分。已经用的是一种叫“上”，另一种叫“下”。这种叫法并无含义，物理学家若愿意，也同样可以把一种夸克叫“艾丽丝”，另一种叫“阿尔伯特”。上、下夸克所带的电荷分别为 2/3 和 $-1/3$，质子由两个上夸克和一个下夸克组成，故总电荷为 $+1$，而中子由两个下夸克和一个上夸克组成，故总电荷为零。π 子则被“解释”为由夸克对组成，即总有一个夸克和一个反夸克配对。一个上夸克加一个反下夸克就成了 π^+，一个下夸克加一个反上夸克就成了 π^-，而一个上夸克加一个反上夸克以及一个下夸克加一个反下夸克都成为 π^0。

到此为止的这一切似乎只是一个备忘录，一种对基本粒子的记忆术。但当盖尔曼和茨威格引入第三种夸克即“奇异”夸克来说明粒子的奇异性时，这种记忆术的威力就显示出来了。通过用奇异夸克来取代一个、两个或三个正常夸克，就可以成功地构造出奇异数为 -1，-2 或 -3 的粒子(用这些负数是历史上定义奇异性时的一个偶然)。质子和中子的奇异数为零，因为它们不含奇异夸克；Ω^- 的奇异数为 -3 是因为它由三个奇异夸克组成，如此等等。八正法的全部结果都表现为这三种夸克或夸克—反夸克对的自然组合。再赋给每种夸克一个确定的质量，其中奇异夸克比其他两种约重出一半，那么所有已知

粒子的质量都能正确地得出来。但是夸克模型真有任何物理意义吗?

盖尔曼是从乔伊斯(James Joyce)的小说《芬尼根彻夜祭》里的一句话[17]里拈出“夸克”这个名称的，甚至他自己在开始时口气也颇委婉。他在引入夸克概念的文章中写道:

> 将夸克看作质量有限的物理粒子(而不是具有无穷大质量极限的纯数学实体)而推测它们的行为方式是很可笑的……在最高能量的加速器上寻找带有-1/3电荷或+2/3电荷的稳定夸克，抑或是带有-2/3电荷或+1/3电荷或+4/3电荷的稳定的双夸克，将促使我们确信并不存在真正的夸克。[18]

是盖尔曼自己真的相信夸克的真实性，却故意把这个概念弄得像个逗乐的把戏那样偷偷塞进物理文献吗?还是如上面那些话所暗示的，他对整个这一套都有怀疑呢?无可怀疑的是，茨威格是认真的;同样无可怀疑的是，他认真的结果是得到的赞扬很少而受到的贬责却甚多。

茨威格1937年生于莫斯科，50年代移居美国，1959年在密歇根大学获数学学士学位，然后到加州理工学院开始研究生涯。为高能质子同步稳相加速器上的一个实验苦干了三年之后，他决定专攻理论，于是在费恩曼的指导下开始研究物理学家对物质世界本质的认识。作为一名新手，他或许少了点年长者的小心和老练，当他意识到介子和重子的八正法分类可以用二或三种亚粒子的组合来解释时，立即就把这些亚粒子作为真实存在来描述，称之为“爱司”*。这种横冲直撞

---

* 原文“Ace”是“王牌”的意思。——译者

看来使他的上级(不包括费恩曼)惊恐，与惊恐相混合的则是这个被他们看作幼稚和不现实的做法的成功。1963 年，茨威格得到奖学金去欧洲核子研究中心学习一年，在那里把自己的研究写成报告以 CERN “预印本”的形式发表，他这样概括：“鉴于我们探讨问题的这种极不成熟的方式，我们得到的结果看来是有点难以理解的。”[19]但这真是他自己的观点吗？还是有别的什么人把不成熟这种评论塞到他的嘴里，或者笔下？他的报告能在 1964 年印出来本身就很不容易。学生茨威格把初稿交给他在欧洲核子研究中心的上级时立即就被退回，他在 1981 年的加州理工学院出版物上回忆道：

> 要想使 CERN 报告以我所希望的形式发表太困难了，我最后只得停止了努力。当某所一流大学的物理系考虑给我一个职位时，该系的资深理论家，也是整个理论物理界最受尊敬的代言人之一，在教员会议上情绪激昂地批评爱司模型是一个“江湖医生”的工作，从而阻止了对我的委任。[20]

不管那位理论家在 1964 年是怎么想的，夸克模型至少提供了便利的方法来计算强子应有的行为。随着最新一代粒子加速器的建成，实验家有了检验夸克假说的手段，他们可以拿电子来轰击质子，用的能量高到应该可以把质子内的夸克分散开来。用加利福尼亚州斯坦福 2 英里(3.2 千米)长的加速器(斯坦福直线加速器，简称 SLAC)做了这种实验，电子被加速到具有 20GeV 以上的能量。电子从靶内的质子上散射的方式清楚地表明，在质子内部有着带电荷的质—能集中区域，正像许多年前的卢瑟福实验证明在原子内有一个集中的核那样。在 20 世纪 60 年代后期的大约同一时间，在欧洲核子研究中心作了用

中微子束而不是电子束来探索质子的实验，结果显示在质子内部必定还有电中性的“物质”。但是，无论对质子的轰击是多么猛烈，也无论用来轰击的是什么粒子，当时和后来的实验都证明，那假定存在的夸克一个都不可能被撞出来。

对强子中联系着夸克的中性物质的解释原则上是简单的，尽管它提出了关于何种基本理论能被构造出来以解释有关过程的新问题。既然质子和中子聚在一起是由于交换 π 子即强力的传递粒子，那么夸克必定也是以同样方式聚合在一起，这种在夸克间交换的粒子叫做“胶子”，因为是它们把夸克胶合在一起而成为质子、中子等等。乍看起来，这似乎意味着有第五种力要考虑。但目前的认识是，胶力才是自然界真正的“强”力，而核物理中的所谓强作用实际上只是胶力的一个附带效应，这与把原子聚合成分子的电力有少量剩余表现为分子之间的微弱电磁力的情况有些相似。

但在 20 世纪 60 年代后半期夸克模型取得实验进展的同时，也存在着问题。为什么夸克只能是三个一组或作为夸克—反夸克对出现呢？最深刻的难题，倒有点挖苦意味地是标志着八正法登峰造极胜利的 Ω^- 粒子，以及那些有一种与它相同的重要性质的粒子。按照夸克模型，Ω^- 是由三个奇异夸克组成的粒子。但这三个夸克的自旋必须相同，也就是说它们是在同一状态。与此相似，实验发现了一种可由三个自旋相同的上夸克的组合来最好地解释的粒子，还发现了由三个自旋相同的下夸克组成的粒子。但夸克是费米子，泡利不相容原理说不能有两个费米子是在同样状态，更别说三个了。有可能是夸克不服从不相容原理吗？或是有什么别的办法能使 Ω^- 粒子里的三个夸克(还有上述其他粒子里的那三个“同样”的夸克)得以相互区分呢？

一个好的夸克理论必须回答这些问题以及其他更多的问题。符合

需要的“好理论”原来就是场论。场论的复兴在 20 世纪 70 年代导致了一个好的夸克理论，在 80 年代又带来了建立一个所有场的统一理论的希望。但这次复兴是由于对轻子和光子而不是对强子的研究中的突破，标志这一突破的是一个把电磁力和弱力结合在一起描述的新理论，即电弱理论。不过在查看这个新理论是怎么建立又怎么帮助理论家去寻找更好的强力模型之前，需要再次探究一下数学家的宝库，把他们最有用的概念工具再取出一件来。

规范事物的性质

按照新的图像，普通物质的相互作用全都可以归结为是在四种粒子之间发生，即上、下夸克和电子及其中微子。对于中子衰变成质子并射出一个电子和一个反中微子的过程，夸克理论说的是，中子里的一个下夸克变成了上夸克并射出一个 W^- 粒子，后者又变成电子和反中微子。对这个反应也可以有另外一种看法，即下夸克把一个虚 W^- 粒子给中微子，使中微子变成电子，而下夸克自己则变成上夸克。电子和它的中微子是与强子世界里上下夸克对应的轻子。这些相互作用全都可以用像图 7.3 那样的散射图来描绘；数学上讲，一个粒子的时间进程与它的配对反粒子的时间倒流是一回事，所以一个基本的图可以对任一个或是所有的基本相互作用都有效。

历史上，对弱力认识的进展是开始于夸克概念被提出之前，所以那些方程式和图以前是、现在也常常是用质子和中子而不是上下夸克来表述的。这不会影响所作论证的效力，而我将交替使用这两种表述。但应当记住，由于只涉及普通形式的物质，它们组成了太阳和恒星、星系、星际介质、行星以及我们自己，在相应水平的描述里就只需要很有限种数的基本粒子，仅仅四种。如果实际存在的粒子仅此四

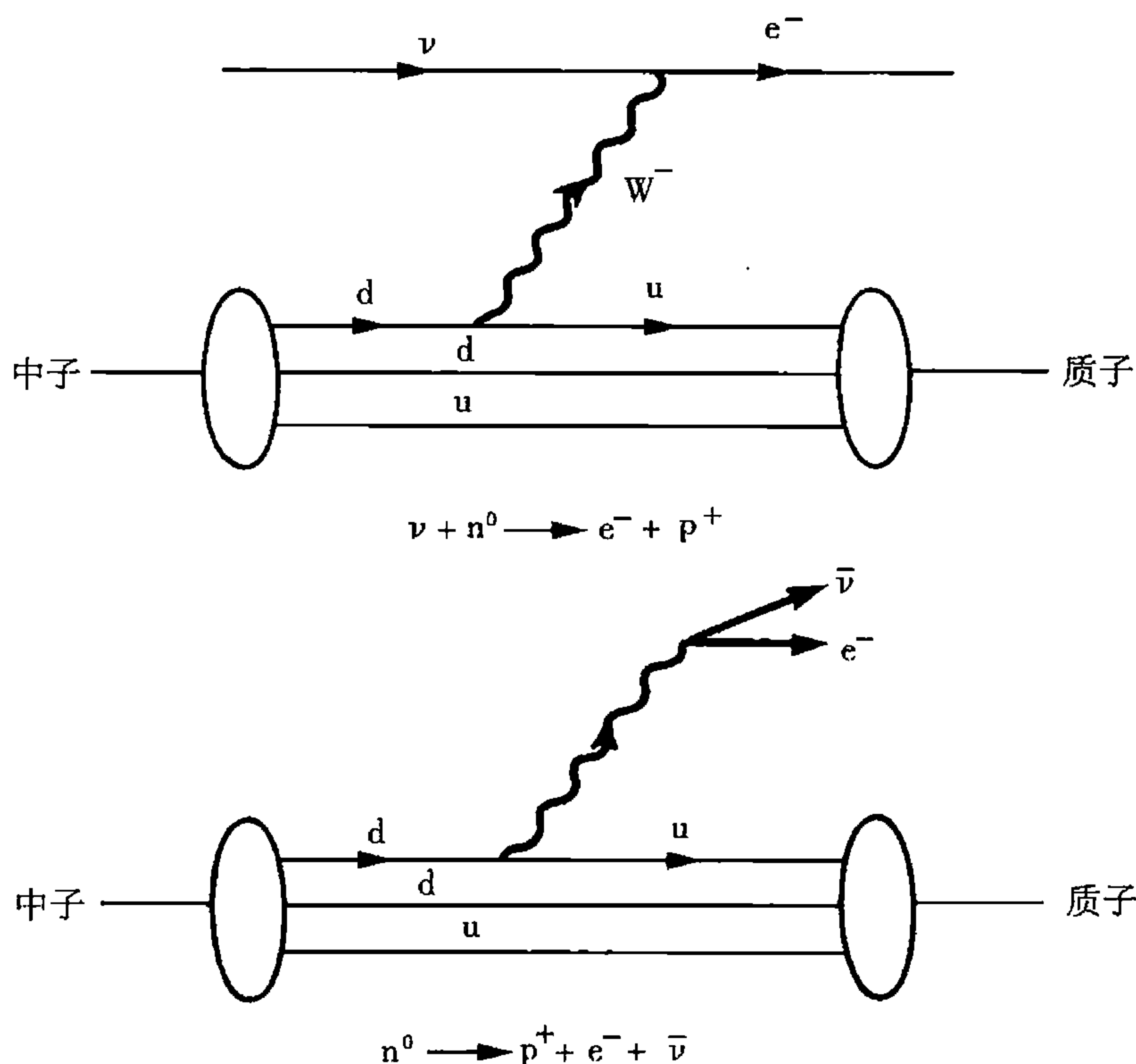


图 7.2　粒子过程的现代图像显示，中子通过弱作用衰变成质子的过程，在更基本的层次上看是一个下夸克变成了一个上夸克。这个过程中释放的 W^- 粒子能与一个中微子作用而生成一个电子(上图)，或者完全等价地说成是 W^- 粒子自己衰变为一个电子和一个反中微子(下图)。

种，则本书迄今所讲的几乎全部物理和宇宙的演化，就会与这个水平的描述一样。

物理学家在 20 世纪 50 年代试图对弱作用构造一个更完整的场论时，很自然地会先查看当时已有的场论，即引力场尤其是电磁场的理论，以确定一个“好”的理论应该具有何种性质。能用于描述这些场的最有力的概念之一是对称性。例如，电场的对称性表现在带电粒子之间的力。如果把一些带电粒子排列起来，其中有的带正电有的带负电，并测量它们之间的所有作用力，然后把每个电荷都变得相反，正的变成负的，负的变成正的，而保持它们的位置不变，那么作用在每个粒子上的力都与原先的完全一样。这样的对称性叫做整体对称性，

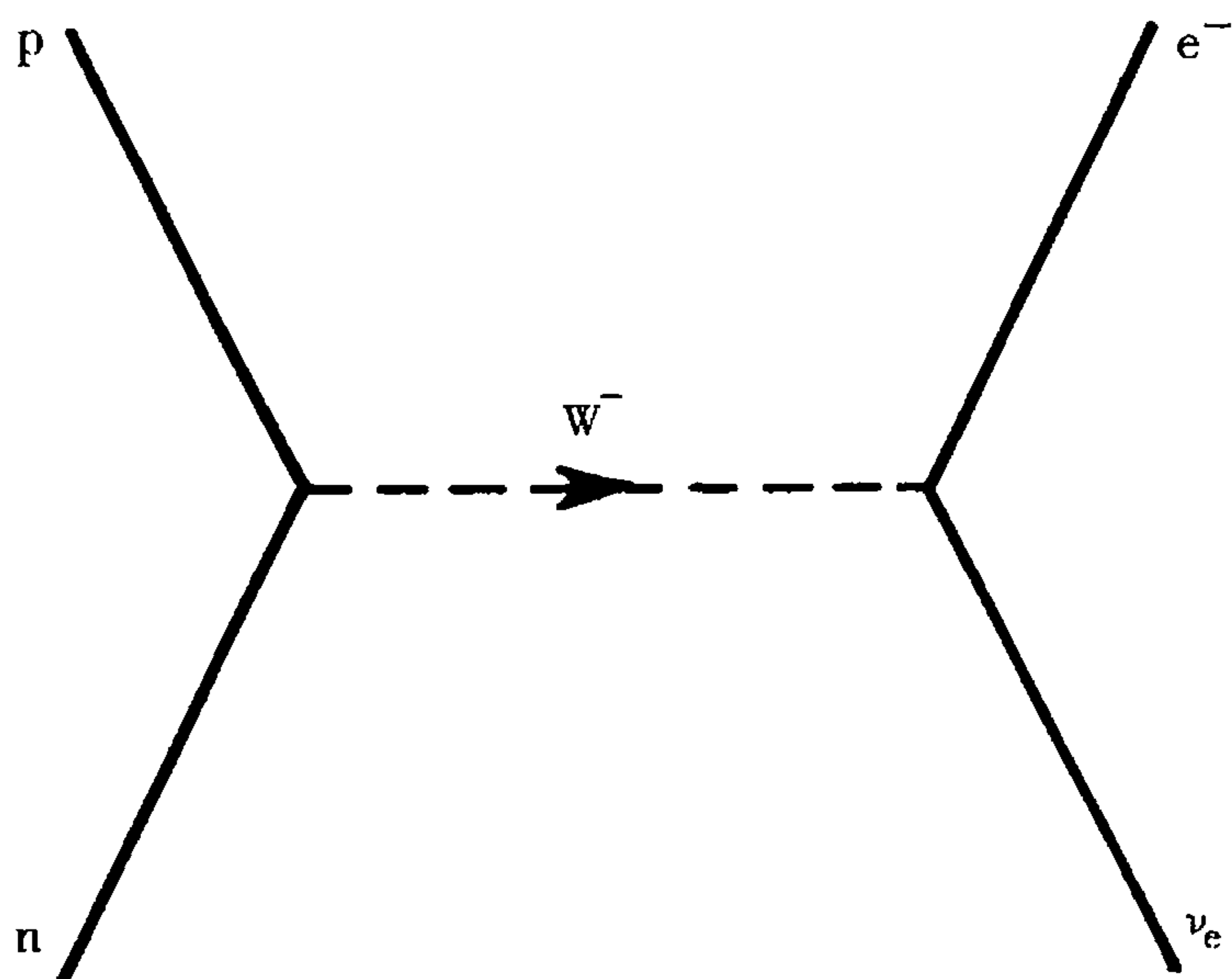


图 7.3 所有的基本粒子相互作用，如图 7.2 所示的作用，都能用像本图这样的散射图来描绘。

每个电荷(严格说来是宇宙中的每个电荷)都必须同时反号，才能保持原来的力场不变。

其他的物理定律和粒子性质也能用对称性来描述。正电和负电可以被认为是镜像，是某种基本“状态”的相反形式。如果不管电磁力而只看质子和中子的其余性质，它们就很相似，相似到使物理学家把它们看作一个基本实体的两种可能的“态”，这个基本实体就叫做核子。电荷的问题已放在一边，那么由什么来决定一个核子究竟是质子还是中子呢? 正如“正”和“负”被用来表述电荷的不同形式那样，也正如夸克被任意地给以“上”和“下”这样的名称那样，物理学家也给使中子和质子相区分的性质起了名。他们把这种每个核子都具有的性质叫做同位旋，它像一个箭头一样，只能是上下向或横向。但它不是在日常生活的三维空间里取向，而是在某种表示核子内部结构的抽象的数学空间里取向。

不妨想象把所有核子的同位旋都同时改变，从而宇宙中的每个质子都变成了一个中子，每个中子都变成了一个质子。这等于是把每个核子同位旋的取向都转动 90°，即转一个直角。重要之处在于，强力不会受这种变换影响，正如电力不受所有电荷反号影响一样。在两种核子态之间即质子与中子之间，或者在更深的层次上讲是在上、下夸克之间，有着一种基本的对称性。所以，当单个中子真的变成了质子时，那个特定粒子的局域对称性就被扰乱了，就有了一个*局域*的对称性变换。但是物理定律应该保持同一，应该与宇宙中所有质子和所有中子都转换了角色时一样。宇宙是怎样记录局域对称性变换的呢？在现在情况下是通过强力本身。自然界的基本力于是就同基本对称性深深地联系在一起了——不仅有整体的而且有局域的对称性变化。



图 7.4　中子与质子的差别可以用核子神秘的“内部箭头”的取向来表示。这个箭头就叫同位旋。

对称性变化能以许多方式发生，但是从数学上讲，隐藏在物理定律中的对称性原来是最简单类型的。它们被称作规范对称性，并且也有局域对称性，这种局域对称性限制了它们的性质并使得计算它们的效应成为可能。

“规范”一词只是数学家用来描述场的一种性质的标记。首先在这个意义上使用这个词的是德国数学家外尔(Hermann Weyl)，时间是在第一次世界大战后不久，那时他在试图建立一个把电磁力(麦克斯韦方程组)和引力(广义相对论)统一起来的理论。规范变换是指某

物理量在所有各处的值同时变化，如果一个场在这样的变化后保持不变，就说它具有规范对称性。前面刚讲过的那个想象的电荷系统就是一个好例子。如果真的在实验室里设置这样一个系统并测量电荷之间的所有的力，就会发现在把整个实验室充电到高电势时那些力保持不变。[21]只有电势之间的差别才是重要的，这就是老鼠能在地铁通电的铁轨上跑得欢的道理。整个老鼠是在同一个电势上，没有电流在它身上流过。但若一个人用手去碰铁轨而身体的另外某部分却与地面接触，那就有问题了，因为他体内就有电势差从而有电流流过。

所有带电粒子的电势都同时增加同样的量时，粒子间的电力保持不变。这种规范不变性是另一种类型的对称性，是引力场也具有的。但若只是电荷系统中的一部分电势升高又会怎样呢？这时会有电流出现，正像一个人或一个物体的一部分碰到通电铁轨上时一样。运动电荷产生出一个新的场，即磁场，它可以用与电势类似的磁势来描述。磁场恢复了描述该系统的方程的对称性。如果想象把实验室中或宇宙中的电势作了复杂的改变，这里升高那里降低，那么总可以通过磁势的补偿改变来抵消电势改变的影响。电磁理论，即包括了电和磁两者的理论，因而是在局域规范变换下不变的。麦克斯韦方程组描述的正是这种最简单类型的场，它既遵守这种对称不变性，也服从狭义相对论方程。

这种类型的对称性深刻地联系着广义相对论的等效原理。爱因斯坦教导我们，加速度总可以由引力来抵消。加速度代表力。如牛顿教给我们的，力等于质量乘以加速度。在一个以恒定速度运行于太空中的实验室里，从实验室的这一端到另一端没有引力势的改变，这个情形就与设置在同一基线电势上的电荷系统一样。在这样一个实验里做的实验完全服从牛顿定律。这些实验所显示的对称性也与那个电荷系

统所显示的对称性相似。而在地面上，由于地球的引力，实验室的顶部和底部之间就有一个引力势差。这就相当于把那个电荷系统实验室一端的电势弄得比另一端高。对称性也就不复存在了。

再回到那个想象的空中实验室，现在改为不时从它里面发射出火箭来推动它，于是实验室里粒子运动的轨道就会改变，就像有神秘的力在起作用。那种力精确地等价于引力。例如，要使这个远离任何大质量的实验室做圆周运动，就得给它以持续不断的推动，实验室里的人可以通过测量室内的力来推断出这种圆周运动。但若那个空中实验室就是在环绕地球的轨道上，那么由于它是在做圆周运动而不是直线运动，所"应该"在室内显示出来的那些力，就被来自地球的引力精确地抵消了。实验室的运动就也是一种自由下落运动。原则上，正如能够用磁势来补偿电势的改变一样，即使是由火箭造成的最猛烈的冲击也可以由变化的引力势来抵消。或者换一种方式来说，原则上可以在实验室周围排列一些物质团块(行星、恒星、黑洞，或随便什么)，使得实验室沿一条最古怪的扭动路线在空中飞行，但总是在做自由下落运动，正如在围绕地球的圆形轨道上的那种自由下落运动一样。这个方案难以实施倒无关紧要，重要的是对称性已经引进了方程式中。引力场在局域规范变换下是不变的。

但我们对自然界的力所能"知道"的一切，只是它们如何影响运动，如何在这里使一个电子的轨道偏转，在那里推动一个质子，等等。自然界里其他的力在粒子层次上所起的作用，与引力对整个宇宙所起的作用相同，即都是作为一种手段来抵消掉由局域对称性变换所造成的扰动。在电磁场的量子表述即 QED 里，力等价于带电粒子之间的光子交换。只有在光子是一种自旋为 1 而质量为零的粒子时，粒子和相连的场的变化才能抵消，以保证局域规范对称性。正好具有这

些性质的光子的存在，被今天的物理学家看作是规范对称性的要求，或是说是对规范对称性方法将成为解开自然界奥秘的钥匙的证实。如果把这种方法从 QED 取出而用到对弱场和强场的描述中去，又将会怎样呢?

注　释:

1. 现仍存于伦敦皇家学院的法拉第博物馆。

2. 严格地讲，波函数描述的是正负电子对，下文很快会再谈到。

3. 更精细地讲，即使标量场也可以有方向，因为场处处不同，而与场相作用的物体总是寻求一个最低能量状态。一只球在地球引力场中下落就是这个道理。但是那个带电小物体即使在一个完全均匀的电场中也将沿力线运动，这才是两种场的关键区别。

4. 这个比拟对推斥的情况来说很好，所以我不能不用它。但它对解释带相反电荷的粒子，如电子和质子，如何相互吸引却毫无用处。遗憾的是，量子世界并不能总以这种舒适的比拟来描述。

5. 正因为在核内的是虚粒子而不是真实粒子，所以才能用来把质子和中子粘在一起而又不增加核的质量。在用量子场论处理强相互作用时，也会出现与 QED 里一样的无穷大问题，也是用同样的办法即重正化来对付，这将在下文谈到。但是那将要求对粒子的“推广”比现在所讲的更“基本”一步。

6. 我起初对使用这个术语感到不安，因为它没有顾及粒子“实际上”是场量子这一事实。但是“粒子”和“场”都只是我们为方便而使用的标记。在我同粒子物理学家克洛斯(Frank Close)讨论在我的讲述中怎样才能最好地使用这些标记时，他说这无关紧要。他在看诸如欧洲核子研究中心进行的那些“粒子”实验并试图理解所发生的事情时，所考虑的是在所研究的相互作用中通过的动量流或动量转移。他说，这才是真正紧要的。粒子和场量子都有动量，只要你知道动量往何处去，标记就是次要的了。

7. 一个在磁场中顺时针运动的电子的轨迹，当然与一个逆时针运动的正电子的轨迹是一样的。安德森的成功正在于确定了那些留下轨迹的粒子究竟是沿何方向运动，也正是这种精巧性才使得他的工作值得授予诺贝尔奖。我在这里说“安德森测量了粒子轨迹的弯曲并发现了正电子”倒是很容易，安德森当年那么干时可就艰难得多了。

8. 虽然中子不带电从而反中子也不带电，它们也像另外那两对一样有明显区分，狄拉克的计算的意义对它们也像其他对子一样重要。光子的反粒子就是它自己。这些精妙内容将在下文讲述。

9. 严格地讲，这个分类也应该把反粒子包括进去，但凡适用于质子的也适用于反质子，依此类推。

10. 中性 π 子的反粒子与其自身不可区分，至少对所有观测到的相互作用而言，中性 π 子也是自己的反粒子。光子也是如此，虽然原则上也可以建立描述反光子的方程，但实际上光子与反光子是一回事。

11. 他还发现了第三种形式的辐射即 γ 射线，后来被证认为高能光子。

12．值得指出，不可能把电子设想成是存在于中子“内部”，也就是把中子看成一个电子和一个质子由电磁力抓在一起的合成物。不确定原理连电子限制在原子核直径的范围内都不允许，更不要说在单个中子之内了。要把中子转变成质子和电子，除了别的之外必须依靠的是相对论的质能关系式，这样才能把质量转变成能量然后又转变成别种形式的质量。由 β 衰变所产生的每一个电子都是新创造的粒子。

13．费米最早的模型是让所有的作用都发生在一个点，这等于是给了 W 粒子零尺度和无穷大质量。用有限质量的粒子来表述弱力是 1938 年由瑞典物理学家克莱因(Oscar Klein)开始的。

14．或者如狄拉克一定会说的，是掩盖起来了。

15．*The Eightfold Way*(New York：Benjamin).

16．当然在 20 世纪 60 年代也有与之竞争的理论。直到 1970 年，每年发表的有关高能物理问题的理论文章中，属于我在这里描述的研究路线的只占不到一半。为缩短篇幅，我这里再次只限于跟随研究主流，连夸克理论本身发展过程中的许多回头路、死胡同、反反复复等等都略而不提。

17．“为检阅者马克，叫三声夸克”(Three quarks for Muster Mark)，从上下文看，其发音与“吠叫声”(Bark)而不是与“猪肉”(Pork)押韵。

18．见盖尔曼的论文，刊于 *Physics Letters*，Vol. 8(1964)，p. 214。亦被引用于 Andrew Pickering，*Constructing Quarks* (Edinburgh：Edinburgh University Press，1984)，p. 88。

19．两篇 CERN 预印本均发行于 1964 年，编号为 8182/TH401 和 8419/TH412。更易于理解不过是第二手的材料由 Pickering 的书给出。

20．Isgur(1981)，p. 439。茨威格在同一篇文章里还说：“盖尔曼有一次告诉我，他的第一篇夸克论文是寄给《物理学快报》发表的，因为他肯定《物理评论快报》不会同意发表。”

有趣的是，盖尔曼被授予 1969 年诺贝尔奖是因为他对粒子物理的其他贡献，其中突出的是奇异性和八正法。甚至在 1969 年夸克理论也不被认为是对粒子世界认识的明显进展，在成果清单上的等级仍很低。茨威格还没有得到诺贝尔奖，即使夸克理论现在已经成为认识宇宙的基础，而且他 1964 年的文章比盖尔曼的要完整和详细得多。是茨威格仍在为他年轻时的冒昧付出代价吗？如果不是，或许诺贝尔奖委员会将很快认识到自己的疏忽。

21．这不只是一个假想的“理想实验”，而是能做的并且已经做过。

第八章　寻找超力

用一种数学描述来包括自然界所有的力，这是自爱因斯坦提出引力的场论即广义相对论以来物理学界梦寐以求的目标。20 世纪 20 年代所作的早期尝试是把引力和电磁力统一到一个理论里，那时物理学家手中也只有这两个研究得比较充分的理论。那些尝试都失败了，但在那些构造统一场论的失败尝试中发展出来的某些技巧却在 20 世纪 80 年代后半期的物理学中复活，并被证明非常成功。引力是自然界的四种力中最弱的，又最难与其他三种力相调和。虽然引力的作用范围很大(引力场的粒子即引力子与光子一样是无质量的)因而主宰着宇宙整体，但在各种力都有效的范围内它却很容易被其他任何一种力压倒。把这张不到 1 克重的纸片稳定在我的书桌上，用的是整个地球的质量。但只要我拿塑料笔在羊毛衫上摩擦使笔带上电荷，再把笔伸到纸片上，就能克服整个地球的引力把纸片提起来。是笔作用到纸片上的电力使纸片从桌面上跳起。引力的确是非常弱的力。电力没有主宰宇宙的唯一原因是，几乎在任何地方正电荷和负电荷都相互平衡，于

是就没有净余的电荷去影响远处的恒星和星系。弱核力和强核力也比引力强得多，幸运的是作用范围很小，因为它们是由有质量的粒子(场量子)来传递的。大略地说，强力比电力强1000倍，比弱力强100 000倍(所以电力约比弱力强100倍)。但强力比引力强10^{38}倍，所以建立一个对强、弱、电力的统一描述比把引力和其他任何一种力统一起来要容易得多，也就不足为怪了。

当物理学家认识到有四种而不是两种基本力要对付时，统一场论的问题就比在20世纪20年代时更令人生畏了。有很少几位研究者，其中最著名的是爱因斯坦，埋头苦干想搞出一套统一理论的方程式，把引力、电磁力和其他力都装进一个包里，但即使是爱因斯坦都没有取得什么成功，虽然他把自己生命最后30年的绝大部分时间都花在了统一电磁力和引力的努力中。当成功开始来到时，它却是从与爱因斯坦正相反的一“端”走来的。爱因斯坦是从在大尺度上支配宇宙的引力开始的。但在今天的物理学家看来，把引力加进统一理论是整个难题的最后堡垒。他们步步为营，先从支配原子的力开始，实际上，是从两种强度最接近的力开始，然后再外推到整个宇宙。弱力第一个被给予“正规”的场理论模型，然后加进了电磁力而成为统一的电弱模型。今天又有了强力的“正规”规范理论，还有了表明如何把强力加进电弱场而成为大统一理论(简称GUT)的清楚迹象。现在还没有能统一这三种力的唯一的大统一理论，但是几乎可以肯定包括了这样一个统一理论在内的这类模型的共有性质已经被概括出来了。在20世纪结束之前也确有希望把引力带进圈子里来。

这些终极进展必须包括对极高能量粒子过程的认识，能量高到相应的物质密度远大于原子核物质的密度。所以，向统一场论前进的步伐，实际上也是向创世瞬间回溯的步伐。现有理论能告诉我们宇宙在

创生后的 10^{-35} 秒之前的状态，这已经追溯到了时间概念本身开始具有意义的时刻。这或许就是一个清楚的迹象，表明这些包含了夸克和轻子的理论确实正在到达物理学的某种基本层次。如果这些模型能把我们带回到创世本身，那么再问是否还有更进一步的真相就没有意义了，正如问时间开始“之前”发生了什么是没有意义的一样。

在 20 世纪，认识宇宙和探索大爆炸的每项重大成果，都是随着对很小世界认识的进展而取得的。19 世纪后期和 20 世纪初期发展起来的光谱学和对原子的认识导致了量子物理学在 20 世纪 20 年代的建立，与量子物理学并肩而来的是宇宙距离尺度的发现(本身又依赖于对原子的认识)和红移的发现。随着 20 世纪 30 年代和 40 年代核物理的发展而来的，则是 40 年代伽莫夫稍嫌早产的大爆炸模型，还有 50 年代霍伊尔及其合作者关于核合成的工作，更有 60 年代标准模型的胜利。与此同时，物理学家还进入了对更高能过程的探索，这些过程涉及到核子“内部”的粒子。在 60 年代尤其是 70 年代，这种探索促成了物质概念的更新和统一场论的起步，在 80 年代又转过来正在引出对大爆炸的极早阶段的新认识。如果新理论如迄今所表现的这样能站得住脚，物理学将很快达到其基本目标，其实也是形而上学的基本目标，即用一套方程来描述一切，这意味着对从创世时刻起到时间的终结为止的整个大爆炸过程的认识。一个统一的场，即英国物理学家戴维斯所称的“超力”，乃是不仅认识今日世界如何运作，而且认识它如何成为今日模样的关键。探索超力的成功之路是在 1954 年起步的，那是在爱因斯坦去世前一年，一位在中国出生在美国工作的物理学家同一位美国同事一起发表了一篇文章，把局域规范理论运用到了强力问题上。他们的模型作为对强力的描述并不特别成功，但却标志着一个概念上的突破，从而鼓励了其他研究者用类似的技巧去探讨别

的问题。有趣的是，开拓的是强力的荒原，而首先结出的果实却是对弱作用的更好认识。

电弱统一

杨振宁于1922年出生在中国合肥。他父亲是一位数学教授，他自己曾在中国昆明的大学和清华大学学习并获得学士学位，然后于1945年到芝加哥大学当研究生，1948年获博士学位，导师是特勒(Edward Teller)。此后杨振宁又在芝加哥给费米当了一年助手，于1949年进入普林斯顿高等研究院，一直干到1965年。杨振宁对沿着与QED相同的路线建立强作用的场模型的可能性很有兴趣，从在芝加哥时起到1954年止的期间内断断续续地研究这个问题，但收效不大。然后他离开普林斯顿去了布鲁克黑文国家实验室一年，在那里他与一位理论家米尔斯(Robert Mills)共用一间办公室。

米尔斯和杨振宁一起构造出了一个强作用的规范不变场论。在杨—米尔斯理论里重要的对称性是上文中已提到的同位旋对称性。按照这样一种对核子的描述，质子和中子分别由数学空间里的垂直箭头和水平箭头表示；如果有一个局域对称性，那就意味着在不同时间改变宇宙中不同地方的单个核子的同位旋是允许的。换句话说，存在着使单个质子变成中子或是反向变化的相互作用。比较简单的整体对称性当然就是“仅仅”允许在想象中把所有中子都变成质子，把所有质子都变成中子，并且全都是在同一时间。

正如此类其他理论那样，当场中有局域的改变时对称性得以保持的办法，是加进别的什么东西来抵消那个改变。在杨—米尔斯理论中，只有引进六个矢量场才能使物理定律在同位旋作任意改变时都保持不变。这些场中的两个在数学上等价于通常的电场和磁场，它们合

在一起描述光子即电磁力的传递者。另外四个场分成两对，描述两种与光子相似但带有电荷的新粒子，一种带正电，一种带负电。由该理论所阐述的包括所有这些粒子的相互作用，真是复杂得可怕。

显然，对强作用的这样一种认识至少是不完整的。首先是，任何一种“光子”都没有质量，于是就会有无限大的活动范围，但有鉴于事实上强力是四种基本力中作用范围最小的一种，所以它的传递者应该具有较大的质量。不过，这个模型背后的思想过去和现在都非常令人感兴趣。简单地讲，两个带相反电荷的“光子”可以被想象成像质子和电子那样束缚在一起，组成强场的“原子”。讲得更深一点，对关于四种相互作用的理论后来的发展有重要影响的一个主要发现是，由于带电光子的存在，基本粒子的系列变换的次序会对最后得到的状态造成极为重要的差别。

这听起来太复杂，还是一步一步来。例如，一个电子的状态可以通过吸收或是发射一个光子来改变。电子先吸收一个光子再发射一个光子后的状态，与先发射一个光子再吸收一个光子后的状态是一样的（假定电子的初始状态一样，所有的光子也都一样）。相互作用所发生的次序无关紧要，QED 因而被称作阿贝尔理论。[1]

通常的数也是这样。2×4 与 4×2 一样，$6+7$ 与 $7+6$ 也一样。这样的数被称作可交换的，并可一般地写作 $A\times B=B\times A$。但在量子物理里却一般不是如此，即有 $A\times B\neq B\times A$，这样的变量就被称作不可交换的，或非阿贝尔的。杨—米尔斯理论中的带电“光子”就是这种情况。如果一个强子由于同位旋箭头的局部转动而改变，然后又由于第二次不同的同位旋转动而第二次改变，它的最终状态就取决于两次改变发生的次序。杨—米尔斯理论是一个非阿贝尔局域规范理论，所有的基本场其实也都由非阿贝尔规范理论描述，包括电磁场在

内，它是一个更大的非阿贝尔理论的一部分，这一点即将讲到。

所有这些听起来可能会觉得很深奥很有技巧。但你可以很简单地用正在读的这本书(别的书也行)来演示非阿贝尔变换。把一本书平放在你面前的桌子上，封面朝上，正好是你准备读的样子。现在把书转动 90° 竖立起来，封面正对着你。再把书转动 180°，使书的封底正对着你。把书还原成封面朝上平放，把刚才两次转动次序颠倒重做一遍。先把书转动 180°，它还是平躺在桌子上，但书名倒过来了。再把它向上转 90° 竖立起来，结果成了封面对着你，而且是倒立的。还是同一本书，也具有同样的能量，但却是在不同的状态。你刚刚对书作了两个非阿贝尔变换。

虽然 20 世纪 50 年代中期的理论家也清楚地知道对杨—米尔斯理论只需再稍作加工，但像非阿贝尔变换这样的基本思想却很有趣并促成了新的研究路线，对这些思想的兴趣也足以说明陈述该理论的论文在 1954 年发表是正确的。[2] 对强力的研究一直到 20 世纪 60 年代后期都没有什么进展，关于强力的令人满意的理论是在杨—米尔斯论文发表 20 年后才得到的，那时才认识到夸克是参与此类作用的基本实体，而胶子是真正的强力传递者。但在另一条道路上，杨—米尔斯的思想被先后引入了弱作用理论和电弱理论即统一电磁力和弱力的理论之中。

施温格尔(Julian Schwinger)是一个数学神童。他生于 1918 年，14 岁进入纽约城市学院，后又转到哥伦比亚大学，17 岁时在那里得到学士学位，三年后又获博士学位。他曾在加利福尼亚大学与“原子弹之父”奥本海默(Robert Oppenheimer)一起工作，后来又到芝加哥大学和麻省理工学院干过，1945 年得到哈佛大学的教职。一年后即 28 岁时，他成为那所庄严的学府历来聘任过的最年轻的正教授之

一。施温格尔对 QED 的发展作出了重大贡献，并因此而与费恩曼和东京大学的朝永振一郎(Shin'ichiro Tomonaga)分享了 1965 年的诺贝尔物理学奖。[3]

所以施温格尔具备了理想的背景来接过杨振宁和米尔斯的思想并运用到弱力和电磁力。弱作用的游戏规则稍有不同。例如，在 β 衰变里中子转变成质子，同位旋对称就被破坏了。但在该过程中同时还有一个中微子转变成电子(或者有一个反中微子和一个电子一同产生，这是一回事)，也就是说在轻子世界里有一个与强子世界里的同位旋改变类似的变换。这就导致了“弱同位旋”的概念，即一个与同位旋相似但适用于轻子和强子两者的量子参量。1957 年，施温格尔把杨振宁和米尔斯对强力的非阿贝尔局域规范理论运用到弱力和电磁力。与杨—米尔斯理论一样，他的理论里也有三种矢量玻色子，一种不带电荷而另外两种带电荷。也像杨振宁和米尔斯一样，他认为那不带电的场量子就是光子。但与他们不同的是，施温格尔把两种带电的矢量玻色子看作 W^+ 和 W^-，即弱力的传递者。仍然有质量的问题，不得不多少有些人为地把特定的质量加给 W 粒子。但这个理论虽有明显的缺点，却再次给出了有趣的新思想。它暗示着弱力和电磁力“实际上”有着同样的强度，在某种意义上是对称的，但是这种对称性已经丧失了或者说被破坏了，因为 W 粒子有质量(从而有有限的作用范围)而光子没有(从而作用范围无限大)。

这就引出了场论的两条发展路线。加利福尼亚大学伯克利分校的布卢德曼(Sidney Bludman)注重与杨—米尔斯理论的连接，于 1958 年指出，单独的弱力能用有三种粒子的局域非阿贝尔规范理论来描述，这三种粒子是 W^+、W^- 和另一种矢量玻色子，不带电，记作 Z^0 或 Z。这就把电磁力排除在外，而意味着应该有不包括电荷变化的弱作

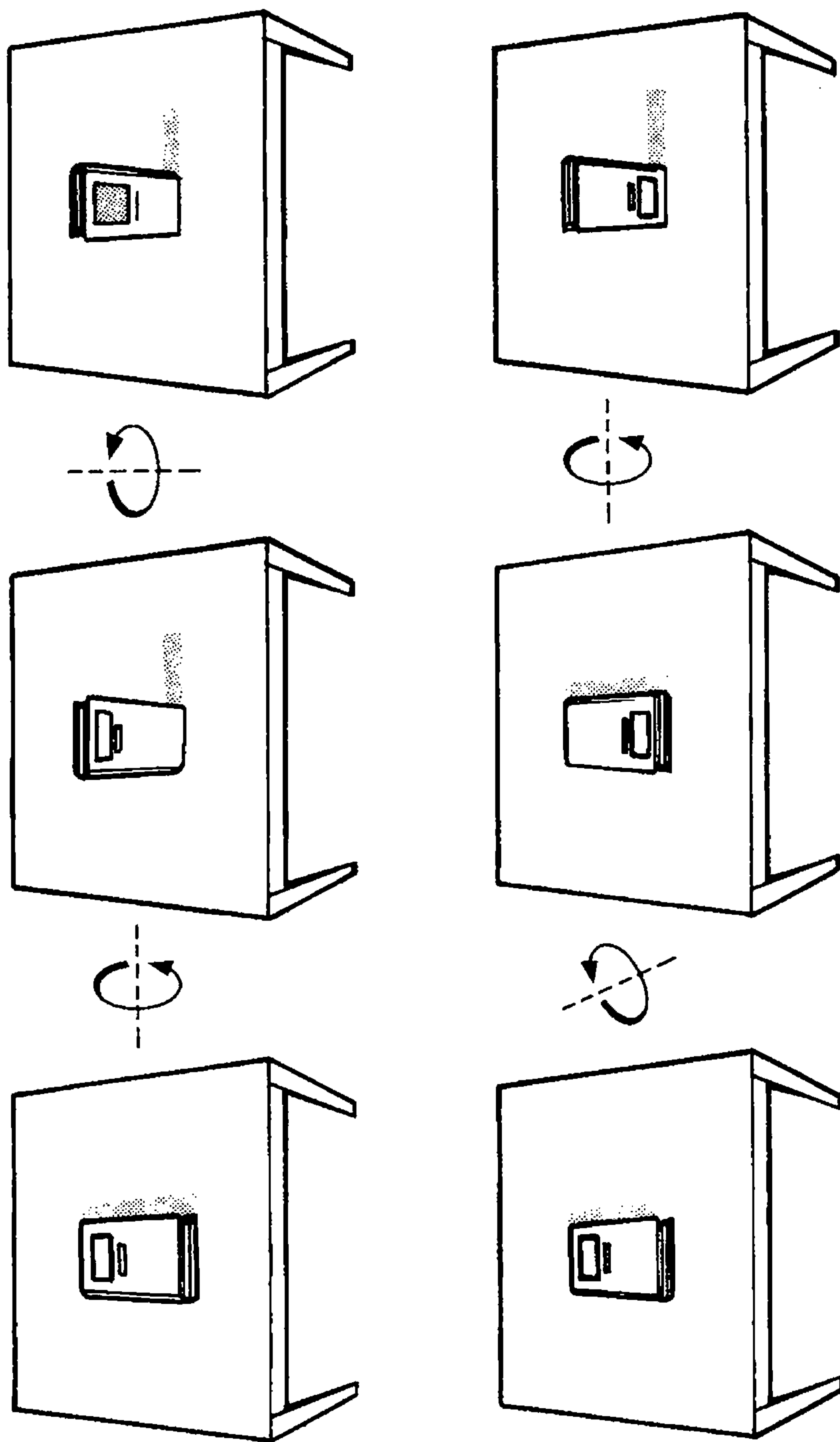


图 8.1　把一本书沿不同方向转动两次，书的最后位置取决于两次转动的先后次序。转动是非阿贝尔的。

用，也就是由 Z 粒子作媒介的作用，称作中性流作用。布卢德曼模型中的所有场量子都是无质量的，所以这个模型是很不现实的。但或许

它比更早已有的模型离“答案”近了一点。

与此同时，格拉肖(Sheldon Lee Glashow)，一位1932年出生于布朗克斯、1954年毕业于康奈尔大学的物理学家，那时正在哈佛由施温格尔指导攻读博士学位。格拉肖想出了一个办法把布卢德曼的模型与对电磁力的描述结合起来，提出了自己的模型并于1961年发表，其中既包含三矢量玻色子组来传递弱力，又有单独一种矢量玻色子来传递电磁力。这种做法立即可见的唯一好处是，那四种矢量玻色子可以混合成为一种质量很大的中性粒子Z，而另一种中性粒子即光子就不再有质量，于是就不会有两种都有质量的中性粒子。但是仍得人为地把质量放进去，以打破基本方程中电磁力与弱力之间的对称性。最糟糕的是，格拉肖理论看来不能重正化而是被无穷大问题所困扰，这种问题也在QED中出现但被适当的数学手法消去。而在早期电弱模型中输入质量的数学手法却不能同时实现重正化。

在同一时间，即从20世纪50年代后期到60年代初期这段时间里，巴基斯坦物理学家萨拉姆(Abdus Salam)和他的同事沃德(John Ward)也在建立一个与格拉肖所提出的很相似的电弱理论。萨拉姆于1926年出生在现属巴基斯坦的江格，在旁遮普大学上学后去了剑桥，1952年在那里获得博士学位，然后在拉合尔和旁遮普大学教书，1954年返回剑桥，在剑桥做的事情之一就是指导其学生肖(Ronald Shaw)的研究。学生所选择的研究课题通常都反映导师的兴趣，肖也不例外。萨拉姆的确对沿着杨—米尔斯理论路线研究自然界基本力规范理论很有兴趣。他于1957年成为伦敦帝国学院的理论物理学教授，还对1964年意大利的里雅斯特国际理论物理中心的建立起了推动作用，那是一个旨在为发展中国家的物理学家提供研究机会的机构。萨拉姆从一开始就担任该中心的主任，此后就有时在的里雅斯特，有时

在帝国学院。

萨拉姆—沃德版本的电弱模型(沃德是一名英国物理学家，20 世纪 60 年代时在美国几所学术机构工作，包括约翰斯·霍普金斯大学)也有与格拉肖版本同样的不足之处，即不得不把质量人为地输入，而这样做的结果是不可能对理论重正化。朝着问题的解决迈出的第一步是在 1967 年，萨拉姆和美国物理学家温伯格各自独立地找到了一条途径，使弱矢量玻色子的质量由方程中自然地(或者说几乎是自然地)出现。这个技巧就是所谓自发对称破缺，它再次是依靠最初在强场情况所提出的概念。

在最弱的场即引力场的情况倒是很容易理解对称破缺。其实我们已经见过了这种破缺。对一个在太空实验室里处于失重状态的宇航员来说没有任何特殊方向。他抛出的笔将沿着被抛的方向漂浮。所有的方向都是平等的，存在着一个基本的对称性。但在地面上就不同了。朝任何方向抛出的笔都同样下落，“下落”是指朝向地球中心。在北极抛出的笔下落，在南极抛出的笔也下落，但这两个“下落”的方向正相反。基本对称性被地球引力场所隐藏，或者说破缺了。

另一种被隐藏的对称性可见于普通的磁针，它总是自动地指着南北方向，尽管电磁学的基本方程是对称的。这种形式的隐对称半个世纪前已由海森伯讨论过，就是那位首先导出不确定关系的海森伯。但最容易理解的例子还是在引力的情况。想象一个完全光滑完全对称的表面，形状就如一顶墨西哥帽子，边缘上翘。如果“帽子”是立在一个水平面上，那么它在地球引力场里是完全对称的。现在想象把一个小圆球放到帽子顶上。只要球不动，一切仍是完美地对称。但谁都知道实际将会发生的是什么。球在那个最高点上的平衡是不稳定的，很快就会滚落下来停在帽檐里。一旦如此，帽子和球作为一个整体就不

再对称了。这个系统中就有了一个特殊方向，即由帽子的中心向外朝着帽檐里的球所在位置的方向。系统现在是稳定的，即处在它能很容易地达到的最低能量状态，但它不再是对称的了。杨—米尔斯类型的

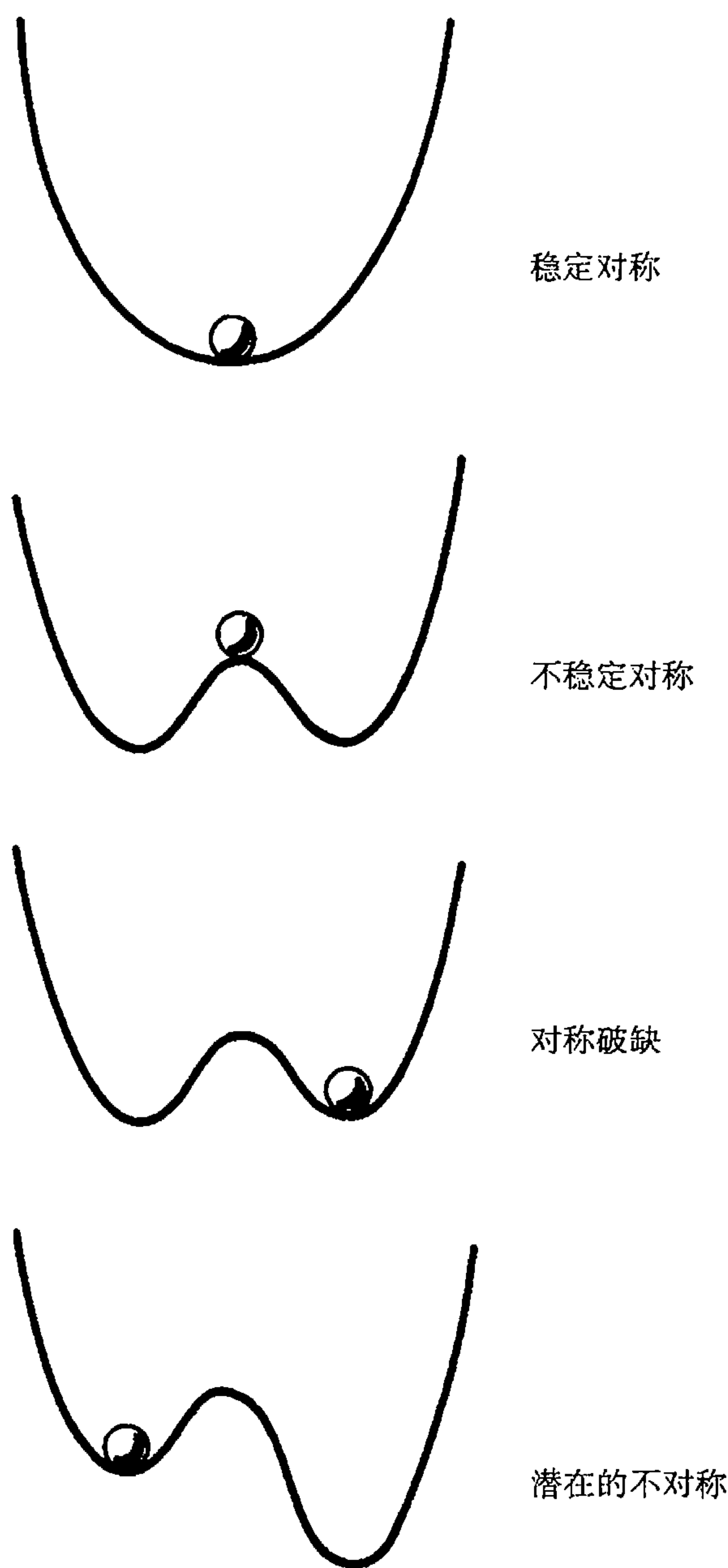


图 8.2　对称破缺可以借助于峡谷中的球来理解。如果只有一个凹处，球就是稳定的，状态是对称的。一旦球有不止一个凹处可选择，情况就复杂了。对称性可以保持，但却是不稳定的；更自然的是稳定态，但对称破缺了。真实世界的事情常常更复杂，还有一种潜在的不对称性，也就是球处在局域稳定态，但不是在最稳定的可能状态，这颇像原子核里的 α 粒子（图 4.6）。

理论中与场量子相连的质量，原来能来自一种类似的对称破缺，发生在同位旋箭头取向的抽象“内部空间”里。

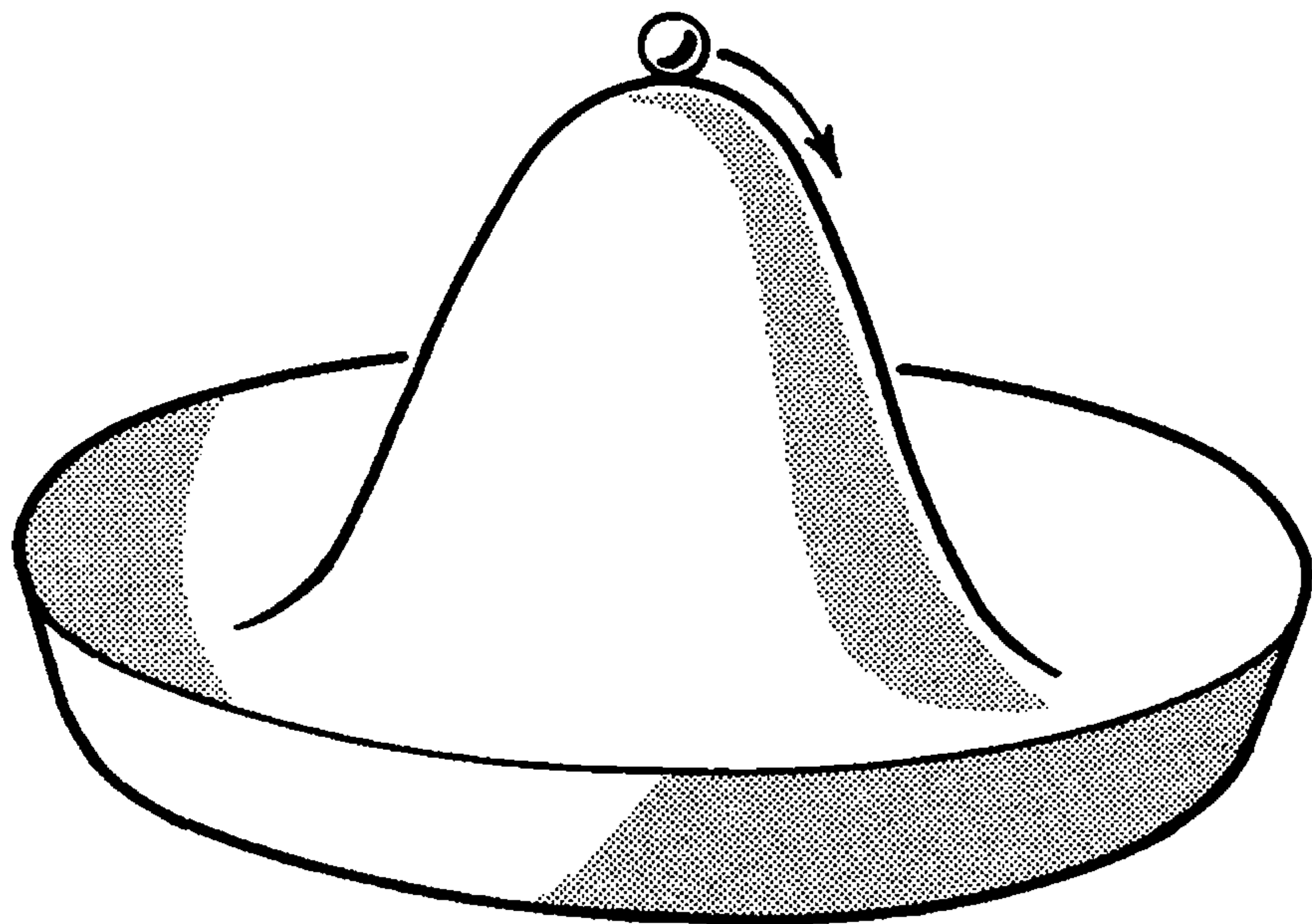


图 8.3 “墨西哥帽”对称性。帽顶上的球表示一种不稳定的对称性，即可以因球沿任何方向的滚落而破缺。

这个思想在 20 世纪 50 年代和 60 年代几位数学物理学家的工作中逐渐酝酿，其成熟则是由于爱丁堡大学的希格斯(Peter Higgs)在1964—1966 年间的研究。希格斯从 1947 年起在伦敦的国王学院学习，1954 年获博士学位，1960 年起在爱丁堡任教。他所提出的机制背后的思路过于复杂，不能在这里细述，但这种机制的大意还是能借助于我们正在熟悉的语言来理解的。希格斯提出，必须把另外一种场加到杨—米尔斯模型里，这种场有一个不寻常的性质，即它不是在场的值为零时，而是在场有一个大于零的值时，具有最小的可能能量。电磁场和大多数其他的场在场的值为零时都具有零能量，而所有的场都具有最低能量的状态就是我们所称的真空。如果所有的场都像电磁场一样，那就等于是说在真空态所有的场都为零。但是希格斯场即使

在其最低能量态也有一个非零值，这就给了真空一个在其他情况下将不会具有的特征。把希格斯场减小到零就必须往系统中放进能量。

这有着深刻的意义。对同位旋来说，希格斯场提供了一个参考系，一个方向，相对于它就能测量表明质子或中子的箭头。通过把同位旋箭头的方向与希格斯场所规定的方向作比较，就能区分质子与中子。但当同位旋箭头在一个规范变换中旋转时，希格斯箭头也同样旋转，使得交角保持不变。曾对应着质子的交角现在对应着中子，反之亦然。没有希格斯机制，就根本没有办法来分辨质子与中子的差异，因为就没有一个参照标准来测量它们不同的同位旋。所有能测量的只是同位旋与希格斯箭头之间的交角，而不是同位旋的任何绝对方向。希格斯场能有此作用，即使它本身是一个标量场，在“真实”空间的每一点上都只有量值，而没有任何特别的方向。

所有这些对矢量玻色子的影响是富有戏剧性的。新的场理论要求四种标量希格斯玻色子，而杨—米尔斯模型给出了三种无质量的矢量玻色子。当这两个成分放到一起时，希格斯玻色子中的三种就会与三种矢量玻色子合并，用萨拉姆的生动语言来说就是，每个矢量玻色子“吃”掉一个希格斯粒子。于是矢量玻色子就既得到了质量，也得到了由希格斯玻色子所携带的自旋。理论上就可以预言，不是有三种无质量矢量玻色子和四种希格斯粒子，而是应该有三种可观测的矢量玻色子，每种都有确定的质量，再加上一种标量希格斯玻色子，它也有很大的质量，但质量的精确值不能由理论预言。希格斯场以完全与观测相符的方式使潜在的对称性破缺。以一种附加的未被发现的粒子为代价，质量自然地出现在杨—米尔斯模型的所有这些变种之中。

希格斯本人一直在研究强场。但他的思想却很快被移植到发展电弱理论上。第一个这么干的是温伯格，时在 1967 年。温伯格是格拉

肖在布朗克斯科学高中的同期同学(温伯格小六个月，生于1933年5月)，1950年毕业，又是格拉肖在康奈尔大学的同学，1954年毕业。温伯格此后就沿着一条不同的途径得到了一个与格拉肖的电弱作用描述很相似的模型，但有着因引入了一个希格斯型机制而带来的优越性。1960年他到了伯克利，在那里待到1969年，接着去了麻省理工学院，1973年又去了哈佛。温伯格对电弱统一的探讨主要是自己做的，当然吸收的是与格拉肖和萨拉姆一样的文化，即一样的物理学知识背景。他对弱作用的兴趣开始于在普林斯顿当博士研究生的时候，20世纪60年代他用自己的方法研究一种与希格斯的等价的机制。他的包含了由自发对称破缺来产生矢量玻色子质量的电弱模型，于1967年10月投稿，同年底刊登在《物理评论快报》上。[4]

萨拉姆从帝国学院的一位同事那里听说希格斯机制，是在温伯格的文章投稿前几个月。萨拉姆就把希格斯机制加进了他和沃德提出的电弱模型，从而得出了与温伯格基本一样的模型，其中有质量的自然出现。1967年他在帝国学院就新模型作了一系列讲座，1968年5月又在诺贝尔专题讨论会上作了一个报告，该报告后来在讨论会文集中发表。

格拉肖、萨拉姆和温伯格由于对创立电弱统一理论的贡献而适时地共获诺贝尔物理学奖，这一理论的意义与一个世纪前麦克斯韦电磁统一理论同样重要。[5] 但这个“适时”却一直等到了1979年。即使是理论家们，其大多数也要有一段时间来充分认识温伯格—萨拉姆模型的意义，因为直到1971年荷兰物理学家特霍夫特(Gerard 't Hooft)*才证明这种电弱理论的确是可重正化的。然后在1973年，欧洲核子

---

* 特霍夫特是1999年诺贝尔物理学奖得主之一。——译者

研究中心的实验得到了该理论预言的那种难以捉摸的中性流作用的证据，那种作用是由中性的Z粒子来传递的。正是特霍夫特的重正化规范理论导致了场论在20世纪70年代的突飞猛进，导致了一个强作用理论，现在又导致了对宇宙自身最早时刻的认识。

规范理论的成年

我这样来讲述的20世纪50年代和60年代规范理论的发展，似乎是很合逻辑很有秩序的不可抗拒的科学进程。其实只是在很有限的意义上才是如此。60年代温伯格、萨拉姆等人所遵循的途径，在当时只是一条偏僻小道。那些涉猎非阿贝尔局域规范理论这种事的理论家，他们作为数学家并不亚于作为物理学家，他们对方程式和对称性本身的兴趣也不亚于对那些东西与真实世界的联系(如果有的话)的兴趣。我们只是凭20世纪80年代后期的事后之见，方知道当时那五花八门的流派里只有这一支才特别重要，才是在导出重大的成果。这能由温伯格论文的遭遇来很清楚地表明，那篇关于电弱统一的论文在1967年发表后的四年之中几乎完全无人理睬。

像温伯格论文这样发表在重要刊物上的学术论文的遭遇，是由一种名为《科学引用索引》(简称SCI)的出版物来监测的，它每年都列出该论文被重要刊物上发表的其他论文所引用的次数。[6] 在1967、1968、1969这三年，温伯格的论文无人引用(甚至温伯格自己也没引用过)。1970年有1次引用，1971年有4次，1972年64次，1973年则是162次。[7]1971年后的这种突然增长完全是由于特霍夫特作出的突破，是他证明了，普遍地讲规范理论，特别地讲电弱理论，是可重正化的。

达到这一成就的过程是缓慢和痛苦的，这里也没有必要去重提那

些弯路和死胡同。故事于是就再次显得是直截了当、并不复杂，但要再次记住，这只是因为我们的事后高明。

故事应该从另一位荷兰物理学家费尔特曼（Martin Veltman）*的工作开始。费尔特曼生于1931年，在乌德勒支大学学习，在欧洲核子研究中心干了5年后又回到母校当物理学教授。他自己沿一条迂回路径搞出了一组与杨—米尔斯场模型等价的规范方程，后来虽然由于1966年与费恩曼的一次讨论而被搞乱了思想，因为费恩曼主张对粒子物理问题作不同的探索，但他最后还是决定接受一位在欧洲核子研究中心工作的英国物理学家贝尔（John Bell）的建议，即最好的前进方向还是建立一个弱作用的杨—米尔斯模型。他以自己的方式着手研究这个课题，用的是路径积分方法，该方法是费恩曼首创的，但在当时（其实现在也如此）几乎没有什么物理学家认真地把它看作一种实用的工具。

所有杨—米尔斯类型的模型明显的主要问题是会出现无穷大并且不能消除。在20世纪60年代中期看来，没有什么办法能去掉这些无穷大，也就是说那些理论原则上是不可重正化的。但随着60年代的流逝，电子计算机对此类研究的作用也变得越来越重要了，费尔特曼正是用计算机找到了能将许多个无穷大消除掉的途径，并且证明原则上可以最终使那些理论完全重正化。他用了许多年进行大量的基础性工作，却没有达到重正化的目标，而是留待下一个人来接过火炬。

特霍夫特于1946年出生在荷兰。[8] 他1964年入读乌德勒支大学，1969年起在费尔特曼指导下攻读博士学位。特霍夫特所选择的课题和所选择的研究方式，都远离科学主流。首先，他感兴趣的规范

* 费尔特曼与特霍夫特分享了1999年诺贝尔物理学奖。——译者

理论，已完全不再时髦。其次，遵照费尔特曼的指教，他决定用费恩曼的路径积分方法来对付规范理论问题。他接过了费尔特曼的许多技巧，得以在1971年发表的一篇文章中证明，无质量的规范理论确实可重正化。这对于一个刚开始做研究的学生来说是优秀的成果，但真正重要的问题当然还是使包含了有质量粒子的理论重正化，W子和Z子，即被认为是传递弱力的中介矢量玻色子，就是有质量的粒子。很久以后，费尔特曼把他在1971年初同特霍夫特的一次谈话讲给安德鲁·皮克林(Andrew Pickering)听，那次谈话是如此深刻地烙在他的记忆里，以致10年多后他还能差不多逐字逐句地重复出来，其文大致如下：

> 费尔特曼：我不在乎做什么和怎么做，但是我们必须至少有一个可重正化的理论，其中包含有质量有电荷的矢量玻色子，至于它看上去是否自然倒没有关系，细节可以在以后确定。
>
> 特霍夫特：我能做到。
>
> 费尔特曼：你说什么？
>
> 特霍夫特：我能做到。
>
> 费尔特曼：写下来，我们就能看出来了。[9]

特霍夫特真的写下来了，费尔特曼也看到自己的学生的确解开了难题。那篇文章于1971年底发表在《核物理》杂志上(B35卷167页)，特霍夫特也于1972年3月被授予博士学位。那时粒子物理学已经在变革，规范理论的中心地位也已经在恢复，这倒要感谢一位名叫本杰明·李(Benjamin Lee)的物理学家，是他使一名冷僻的学生用一种冷僻的方法解决了一个冷僻的问题这一消息在美国传开。李在

1971 年夏天访问了乌德勒支大学，回美国时带走了特霍夫特那年写的两篇文章。李既证实了特霍夫特工作的正确性，又把它用更常规的数学语言写在自己的一篇文章里于 1972 年发表。李的文章使温伯格等理论家重视特霍夫特的工作，并使他们相信，包含了对称破缺和通过希格斯机制引入质量的电弱作用规范理论确实是可重正化的。规范理论终于成年了，至少在理论家所关心的范围内是如此。

1973 年在欧洲核子研究中心作的用高能中微子流轰击一个名为"巨人"的巨大气泡室的实验，给出了包含那难以捉摸的 Z 粒子的相互作用的证据。气泡室里留下的踪迹显示，一个反中微子或中微子恰如电弱理论预言的那样与一个电子发生作用，以 Z^0 粒子为作用的媒介。后继的实验证实了这种解释的合理性，物理学家在分析了对"巨人"里发生的事件拍摄的 300 万张照片之后，找到了 166 个能用中性流来最好地解释的反应事例。实验家们终于也相信，电弱规范理论是有轻子和光子参与的相互作用的最好理论。

这些发现与特霍夫特对电弱理论的重正化合在一起产生了重大的影响，温伯格、萨拉姆和格拉肖因而获得了 1979 年的诺贝尔奖，尽管那时还没有关于 W 粒子和 Z 粒子存在的直接证据。但理论所预言的已不止是这些粒子的存在，而且还有它们必定具有的质量。W 粒子的质量是约 92GeV(稍小于质子质量的 100 倍)，而 Z^0 粒子的质量约为 82GeV。要造出这样的粒子并观察它们的衰变，就需要能够把至少这么大的能量加到碰撞中的加速器里。欧洲核子研究中心在日内瓦建造的正是这样一种加速器，用于使质子束与反质子束迎头对撞。到 1983 年初时，这台加速器已经给出了关于 W 和 Z 粒子存在的清楚证据，粒子的质量也很接近于预言值，它们是在碰撞中产生的，然后就衰变成高能电子和其他粒子。[10]这个发现无疑证实了 1979 年诺贝

尔奖的授予是正确的，使得评奖委员会深感欣慰，于是赶快又把 1984 年的物理学奖授给了该实验组的负责人鲁比亚(Carlo Rubbia)。

这些质量对大爆炸中力的统一的意义是显而易见的。当宇宙的能量密度(温度)足够高时，质量接近于 100GeV 的粒子能以粒子—反粒子对的形式自动出现。但它们并不是作为弱作用的传递者而只能存在由不确定原理所允许的短暂时间，周围自由能量的供给使这些粒子由虚变实，并有着较长的寿命。只要粒子质量小于可得到的能量，粒子就可以像光子那样长久存在，光子与 W 粒子和 Z 粒子之间也就没有区别。能量足够高时，电磁力与弱力之间是没有区别的。区别的出现只是因为宇宙冷却了，对称性破缺了。W 子和 Z 子开始冻结出来是在宇宙温度降到 10^{15}K 即 $t=0$ 之后约 10^{-9} 秒时。那也就是电磁力和弱力开始分道扬镳的时候，一直到人类在日内瓦附近一台机器的极小空间范围里极短暂地重新造出了宇宙在创生后 10^{-9} 秒时到处存在的条件，才把它们再次调和到一起。

1985 年，欧洲核子研究中心的质子—反质子对撞机的能量达到 900GeV，这是一个新的世界纪录，使得制造那些中介矢量玻色子可望成为几乎是例行公事。但是那些由更高级的自然力统一理论所要求的粒子就不同了，没有指望能同样成功地把它们也制造出来。建立在电弱规范理论的成功基础上的新理论指出，那些粒子的质量远远超出任何可想象的人造加速器的能量范围。唯一可以得到这样高能量的场所是早期的大爆炸。所以宇宙就成了粒子物理最新理论的检验基地，正如宇宙学家以前由于大爆炸的新理论而转到粒子物理上一样。我们已经几乎准备好了再继续去探索大爆炸，还需补充的是有关强力理论的一些知识。

带色的夸克

20 世纪 60 年代中期已知的轻子有两组，每组都有一种类电子粒子和一种中微子，即电子与相应的中微子配对，μ 子也与相应的中微子配对。在夸克的概念被引入时，只需要三种就能解释所有已知的粒子。上夸克和下夸克组成一对，而奇异夸克却是孤独的。其实盖尔曼在他提出夸克概念的那篇文章里曾推测过，可能有第四种夸克与奇异夸克配对，组成两对夸克以与两对轻子相称。但这个想法很快被放弃了，因为没有包含那种夸克的粒子存在的证据。而三个看似相同的夸克怎样共存于同一种状态而组成像 Ω^- 这样的粒子，这个问题则紧迫得多，因而使得 20 世纪 60 年代中后期研究夸克理论的物理学家付出了诸多努力，直至最终得以解决。

马里兰大学的理论家格林伯格(Walter Greenberg)，在 1964 年夸克概念引入时很感高兴，因为这使他几年来一些古怪的场论思想有了实际应用。格林伯格最初的兴趣只是发展场论的数学形式，几乎或完全没有想过实际应用。但他的一个称为“仲统计法”的抽象概念，原来却与夸克问题有关。格林伯格很快就把这个概念用到新的强子模型中，并得到了很有趣的结果。这项研究是很专门性的，但归结起来是提出了可能存在服从仲统计法规则的不同种类的“仲夸克”，而 Ω^- 和其他一些强子里那三个看似相同的夸克实际上可以由一种以前不知道的性质来区分为三种类型。这个思想被两位日本理论家接了过去，一个是芝加哥大学的南部阳一郎(Yoichiro Nambu)，一个是锡拉丘兹大学的韩(M. Y. Han)。他们于 1965 年共同提出了一个格林伯格理论的新版本，比起那美妙的仲统计数学来，它使实验家觉得更清楚，因而也就能为更多的物理学家所接受。

这项工作的基本思想是，每种已知的夸克又分为三种，各有不同

的色。色一词在这里只是一个方便的标记手段，就像“上”和“下”的名称一样。但它能使我们懂得，红上夸克与蓝上夸克是不同的，正如红上夸克与红下夸克也不同那样。数学方程给出了这三种夸克应该如何相互作用，而且它们的作用具有优美的精确性。但方程所给出的东西的实质可以借助于简单的色来领会。例如，Ω^-粒子被认为是由三个奇异夸克组成，三个夸克的自旋都一样，但一“红”一“蓝”一“绿”，于是就可以区分，而不是相同状态的相同粒子。这里的色只是一种记忆手段，一种帮助我们理解的智力拐杖。而数学物理学家又向我们保证，这种由类比而来的想象并不太骗人。

至少现在如此。1965 年时，这被看作只是一种游戏，没有更深的意义。南部和韩的模型包括了更多的三夸克组以图免去对分数电荷的需要，这是有点把水搅浑了，但由于当时几乎没有什么人认真看待夸克这个概念，所以也就没有引起什么轰动。但这个模型的确为夸克的行为提供了新准则，包括对为什么它们只是成仨(组成重子)或成对(组成介子)出现这个难题的解答。南部只制定了一条规则，即“允许”的夸克组合必定是无色的，于是就能解释强子怎么划分成这两个族。他说，每个介子必定是由一个某种特定色的夸克和一个带有对应反色的任何一种反夸克组成。例如，一个红上可以与一个反红上，或一个反红下，或一个反红奇异组合，在这几种情况色与反色都在数学意义上“抵消”。他还指出，另一种达到中性状况的方式是在一个粒子里混有三种色，即红、绿、蓝夸克各一个，每一个都可以是上、下、奇异中的任何一种。三个不同色的反夸克也能达到同样目标。但是比如说，单一个夸克或是四个一组都会带有净色，看来是被禁止的。

到了 1970 年，得到了看来与这个夸克色模型相符的实验结果，



图 8.4　三个夸克组成一个质子，即红上 + 绿下 + 蓝上。一对夸克—反夸克组成一个 π 子，即红上 + 反红反下。

色这个概念才开始得到立足的基础。大约在同一时间，格拉肖和在哈佛的两名同事，伊利奥普洛斯（John Iliopoulos）和马亚尼（Luciano Maiani），重新提出有第四种夸克，格拉肖给它起名为“粲”，用以整理对其他一些令人困惑的实验结果的理论解释。1971 年，盖尔曼和弗里奇（Harald Fritzsch）一起接过了色的概念并开始建立一个描述有这三种粒子参与的相互作用的场论，后者于 1943 年出生在茨维考，现在是慕尼黑马克斯·普朗克物理研究所的物理学教授。1972 年秋，他们俩提出，对强子结构的最好描述是用一种杨—米尔斯型的规范理论，其中三种色夸克的相互作用是以八种胶子的组合为媒介进行的。对称性是更复杂了，数目是更多了，但原理与已获成功的 QED 和电弱力理论是一样的。

基本思想再次能由对称性来理解。现在必须设想每个重子含有三个夸克，每个夸克又有一种选择色的手段，即一个内部取向器，类似同位旋的那个横或竖的取向器，但现在有三种指向，对应着三种色。一个对称的整体规范变换是，把每个取向器都顺时针（比如说）旋转 120°，改变了每个夸克的色却保持物理定律不变。而一个局域的对称规范变换则是，改变一个重子里的一个夸克的取向（即色）而世界仍保持不变。在局域变换下恢复对称性的办法，与以前一样是引进新的场，在现在情况下对应着八种胶子，全都无质量（在该理论的最初形式里是这样）并且自旋为 1，也就是与光子类似的矢量玻色子。

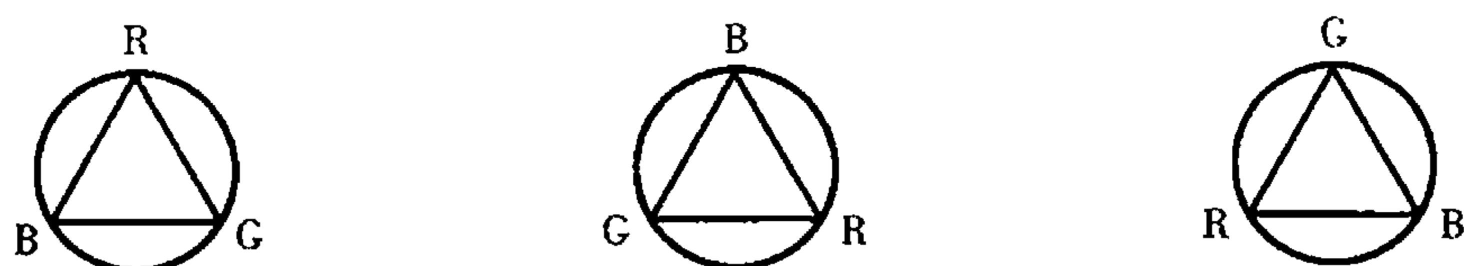


图 8.5　给出夸克的色的内部取向器就像一只旋钮，只能在三个位置间变换，与区分质子和中子的同位旋取向器类似。

这个理论叫做量子色动力学，名称是盖尔曼有意摹拟量子电动力学而取的，现在通常简记为 QCD。它说的是，任何夸克都能独立于其他所有夸克而自由地改变色，它这样做时会射出一个胶子，胶子立即被另一个夸克吸收，后一夸克所发生的色改变正好与前一夸克的改变相抵消，从而保持强子无色。尽管夸克时刻都在像万花筒似地改变色，强子却总是无色的。由于胶子带色，其行为就与光子大不相同。光子不带电荷，相互不发生作用；而胶子甚至在把力从一个夸克携带到另一个的过程中也会相互作用。或许由此而来的最奇怪结果是，虽然“强”力在短距离上(比如说在质子内部)其实是相当弱，但胶子的相互作用却使这个力在更长距离上变强，于是在 10^{-13} 厘米的范围强力就足以克服电斥力而把质子束缚在一起。这就像一根两端各拴有一个夸克的橡皮带，若你不去试图把夸克拉开，橡皮带只是把它们松松地连在一起。你拉得越厉害，橡皮带就伸得越长，它试图把夸克保持在一起的回拉力也越大。这“橡皮带”就是在两个夸克间交换的胶子流。

如果你的拉力足够强，也就是给了碰撞粒子足够的能量，橡皮带最终会断开。但这仍不意味着会有自由夸克显露出来。来自夸克间作用力的能量会在断开的橡皮带两端各产生一个新夸克，这很像一根磁棒弄断后在断开处会“产生”一个新磁极。不会有单个夸克出现，至少总有两个夸克由胶子流即介子连在一起。由于胶子带色，于是也像

夸克一样被迫集体运动，而不能孤立存在，这被认为就是没有探测到孤立胶子的原因。胶子尽管可能是无质量的，却不能像光子那样自己散开。但或许无色的一包胶子（“胶球”）能在诸如欧洲核子研究中心所做的那些实验里显示出来。

70 年代物理学的转折是在 1974 年来到的，斯坦福大学和纽约长岛布鲁克黑文国家实验室的两个小组几乎同时地各自得到了一种新的有质量粒子（现在通常记作 ψ ）存在的证据，那种粒子看来含有第四种夸克即“粲”。这个发现使布鲁克黑文小组的负责人丁肇中和他在斯坦福的竞争对手里克特（Burton Richter）共享了 1976 年的诺贝尔物理学奖。物理学家们对这个发现激动不已，今天还把它称作“11 月革命”，因为实验结果是在 1974 年 11 月发布的。一旦找到了一个粲粒子，实验家们在一定意义上就知道了该到哪里去把其余的找出来，于是很快就发现了粲粒子的整个家族。这族粒子使物理学家有了一个 QCD 的试验台，从而表明 QCD 确实成功地对“新”粒子的行为作了详细的预言。验明了四种夸克和四种轻子，粒子世界看来真是整洁匀称。但是还有一步（最后一步？）要走。

1975 年在斯坦福直线加速器上做的实验显示，可能还有另一种轻子，一种质量为质子 2 倍的“电子”，用希腊字母 τ 标记。一年后在汉堡对此作出了证实。还假定（有很强的间接证据，但尚未证实）存在一种对应的 τ 中微子，这样就共有六种轻子，分为三对。于是理论家就争辩说“应该”还有另外两种夸克，以保持对称性。这两种夸克被起名顶和底，关于底夸克存在的证据于 1977 年得到。对顶夸克的搜寻现在是欧洲核子研究中心的重点工作之一，那里能造出世界上最高能量的碰撞粒子束。这的确应该是路的终点了，因为还有使人不得不信的宇宙学依据表明，宇宙中轻子的种类至多只能有三对或四

对。这些依据是下一章要讲的重要内容。但要记住，假如只有上和下两种夸克，再加上电子及其中微子这两种轻子，宇宙中的几乎一切都会与今天的一样。其余的看来是并无必要地多费了两次劲，它们只是方程式的结果，它们的出现除了没有被禁止出现外并无更好的理由。

表 8.1*

轻子	夸克
e^- ν_e	u(上) d(下)
μ^- ν_μ	c(粲) s(奇异)
τ^- ν_τ	t(顶) b(底)

* 迄今所知宇宙中有三种物质。永久物质仅由两种轻子和两种夸克组成，如表上部所示。由于尚不明了的原因，这个永久世界被对等的、但不稳定而且是更重的粒子复制了不是一次，而是两次。如果只存在电子和它的中微子以及上和下夸克这几种粒子，所有的日常物理和我们周围的世界都不会改变，而粒子物理会容易得多。

电弱理论与 QCD 的结合成功地描述了粒子世界，以至于有时被称作物理学的“标准模型”。但是这种结合仍不完整。QCD 还需要与弱电理论统一成为一个大统一理论，即 GUT；而引力还根本没有包括进去。所以还有许多事情够理论家们忙碌的。[11] 但粒子物理的“实验”却出现了新局面，因为大爆炸提供了检验最新的统一理论的最好场所。

寻找超对称

QCD 尚未能建立得如 40 年前的 QED 那样好。只有用 20 世纪 80 年代的事后高明去回顾时才能指出 60 年代物理学发展的主线，无疑也需要 2000 年或更晚时候的事后高明，才能回顾目前对超力的理论探索的混乱状态并抽出主线。我在这里也不准备把任何一条现在正进

行的研究路线说成是“最好的”或“真实的”，任何一个比如说在1961年曾尝试这样做的人，恐怕都不会把非阿贝尔局域规范理论抽出来作为最好的候选者而加以注视，更不用说那个质子和中子可能是由别的粒子组成的想法了。但我还是在对称和规范不变的基本概念水平上简略介绍一些最有趣的基本思想。在这些思想的基础上可以构成许多详细的理论，我们希望这些理论中能有一个是真正描述了真实世界。如果说最近的情形能给我们什么启发的话，那就是像对称性这样简单而有威力的概念确实能帮助我们区分好的与差的理论。

按照迄今已经解决的和尚未回答的问题来评判，应该说QED是优秀的理论，电弱理论的确很好，而QCD就只能说是还好。这些理论之间的相似性或许倒是最好地标志着理论家正在通往一个更基本的理论的道路上前进，那就是将把自然界的力统一成为一种超力的超理论。电磁场是最简单的，它只需要电荷。弱场的特征则是可取两个值的同位旋，并且联系着两种夸克和两种轻子。夸克则是三个一组并要由一种更复杂的场来描述。但是同样的普遍原理支承着QED的一元性、弱场的二元性和QCD的三元性，并且使得前二者能结合成为一个成功的统一理论。除了是分为三种之外，QCD中的色正与QED中的电荷相似。不带电荷的粒子不能感受电磁场，不带色的粒子即轻子也不能感受QCD的场。

把这些思想沿同一方向继续推进，许多理论家已经在试图构造把电弱理论和QCD压到一个包里的大统一理论。大多数大统一理论都属于同一族，即都追随着由格拉肖和他的哈佛同事格奥尔基(Howard Georgi)在20世纪70年代中期所开创的研究路线。这些理论中都有着五种粒子，例如，其中一种理论中的五粒子族是三种色的反下夸克加上电子及其中微子。粒子族的成员可以互相变换，与质子转变成中

子和夸克的一种色变成另一种色一样，但是现在的旋转取向器有着五个位置，因而就有可能把轻子变成夸克，把夸克变成轻子。GUT描述的是比任何较简单的理论更深刻的对称性，但这是有代价的。

电弱理论需要4种玻色子，即光子、两种W子和Z子。而五维的GUT，数学上称为SU(5)理论，却要求有*24*种玻色子，其中4种是电弱理论中已有的，8种是QCD所要求的胶子，还有12种是“新”粒子，用以传递前所未知的新型相互作用。这些假设的粒子统统用表示未知量的X或Y来称呼。它们能把夸克变成轻子或是变回来，带有1/3或4/3单位的电荷。它们的质量很大，大到在今日宇宙中只能存活极短时间，因而对粒子世界的活动影响甚微。

按照这些理论，三种力(电磁力、弱作用和QCD中的强力)在能量高到10^{15}GeV时都相等，这个能量是使电磁力和弱力相统一的能量的10^{13}倍。与此能量相应的时间是宇宙年龄为10^{-37}秒，温度为10^{29}K之时，这就是说X粒子本身的质量必定是约10^{15}GeV，比欧洲核子研究中心的新质子—反质子对撞机里的碰撞所能达到的最高能量高出10^{12}倍。看不出有什么希望能人工造出这样的条件，这也就是为什么物理学家不得不到大爆炸中去寻找X粒子存在过的证据。但令人吃惊的是，有可能探测它们今天在我们这里存在的一种副效应。

如果质子中的一个夸克能由不确定关系借到足够的能量产生出一个虚X玻色子并把它与另一个夸克交换，那么这两个夸克之一将变成电子(或正电子)。剩下的两个夸克会变成一个π介子，质子也就衰变了。由于X玻色子质量极大，其虚寿命也就极短，因而只有在两个夸克相距不到10^{-29}厘米时它才能从一个夸克转移到另一个，这个距离比质子本身的尺度要小10^{17}倍(即质子尺度的10^{-17}倍)。夸克之间如此靠近必定很罕见，但也还是时而发生，而且发生的可能性是可以计

算出来的。计算结果是，对单个质子而言这种情况要在10^{30}年以上才发生一次，各个理论给出的结果不一样，也有可能至少在10^{32}年里一次都不会发生。宇宙的年龄只有10^{10}年的样子，所以质子仍然到处都是而且看来很稳定也就不奇怪了。但若一个质子在一年里衰变的概率是$1/10^{32}$，那么在10^{32}个质子里就有很好的机会每年能看到一个质子衰变(但不知道是其中哪一个)。

已经设计出了实验来对此作检验，也就是成年累月地观察大量的质子，看其中是否有任何一个发生衰变。1000吨水含有大约10^{33}个质子，而水总是容易得到的。世界上好几个国家正在做这种实验，观测一大池水或一大块铁里的质子衰变。现在还没有什么结论性的证据，这样的或那样的都没有，但多年的努力倾向于把质子寿命的上限设为约10^{32}年，对理论家来说事情就开始变得有趣了。或许某个确定的消息，不管是这样或那样的，将会很快发布。

但总的说来GUT的情况不妙。从规范理论的简单对称性概念出发的研究路线已经变得丑陋而复杂，有玻色子激增的问题，也搞不清重正化的真正意思是什么，每一种吸收进模型中的新力都成了一团乱麻。每次本来只要一个时却有更多的夸克和轻子冒出来，这说明对理论的基本构成成分缺乏一定的制约。令人为难的还有，所有的大统一理论都预言磁单极的存在，而在我们居住的世界里却从来没有发现过。而且，由于可能的规范理论有无限多个，也就难以理解为什么这些特定的理论是在揭示着真实世界。所以，如果摆脱循序渐进的做法，从顶部开始来造一座纸房子，然后再回到根基，那又会怎样呢?

卡尔斯鲁厄大学的威斯(Julian Wess)和加利福尼亚大学伯克利分校的祖米诺(Bruno Zumino)在1974年就是这样做的。GUT使人惊奇的是，把轻子与夸克相联系却仍把玻色子看作不同于物质粒子，而只

是力的传递者。威斯和祖米诺所主张的实际上是，既然对称性是个好主意，为什么不干到底来个超对称性把费米子和玻色子联系起来呢？

停下来对此想一想。费米子与玻色子的区分在量子物理里是件大事。玻色子不服从泡利不相容原理，而费米子服从。它们之间的差别可是比粉笔与奶酪的差别大多了。物质和力真能是同一样东西的两面吗？超对称性说是的，宇宙中的每个费米子都必定有一个配对的玻色子，每个玻色子也必定有一个费米子伙伴。我们在实验中看到的和在日常生活中感觉到的，都只是宇宙的一半。每个夸克作为一个费米子，应该有一个玻色子伙伴，叫做 s 夸克；每个光子作为一个玻色子，也应该有一个费米子伙伴，叫做光微子；如此等等。[12]要解释那些伙伴到哪里去了倒没有什么问题。在这个游戏的初级阶段，理论家们能挥舞手中的数学魔杖，引入某种形式的(尚未确定的)对称破缺来给那些没看见的粒子以很大的质量，在宇宙冷却时把它们撇在一边。主张玻色子与费米子之间存在对称性，这在任何被教育成相信粒子与力的区分的人听来都不能容忍。但真是那么不能容忍吗？我们以前没有经历过类似的事情吗？量子物理说，粒子是波，波也是粒子。对 19 世纪的物理学家如麦克斯韦来说，电子是粒子，而光是波；20 世纪 20 年代的物理学家知道了电子既是粒子又是波，而光子既是波又是粒子。它们又分别是费米子家族和玻色子家族的原型。把波粒二象性推到其合乎逻辑的限度，就应该说粒子波和波粒子是一回事，难道超对称会比这更使我们的日常观念不能容忍吗？只是因为我们已经离开了前两章里所讲的量子物理的根基，并且为了方便而用微小颗粒之间的碰撞和相互作用来描述亚原子世界的事件，超对称才显得像个怪物而使我们感到冲击。如果我们的头脑仍像前面那样能按照量子方程来掌握抽象的概念，因而能恰当地认识量子世界的本质，知道在那里除

非被观测就没有什么是真实的，也根本没有办法说出“粒子”除了相互作用的瞬间以外还在干什么，那么超对称就会显得自然多了。缺陷在于我们的想象而不在于理论。但仅凭很有限的想象也能认识新理论的一个特征，仅此特征就足以使绝大多数其他以“超力”为题的理论大为逊色。超对称(简称为 SUSY)的最激动人心之处在于，使玻色子与费米子相互变换所需要的数学技巧，能自动地和必然地把时空结构也就是引力引进来。

把玻色子转变成费米子的对称性运算与广义相对论的对称性运算在数学上是密切相关的。对一个费米子作超对称变换，就会得到一个对应的玻色子，比如说一个夸克就变成一个 s 夸克。把同样变换再作一次，就会又得到原来的费米子，但是稍稍移动了位置。超对称变换所包含的不仅是玻色子和费米子，而且有时空本身。而按照广义相对论，引力简单地就是时空几何的反映。

于是自然就有一帮有志者接纳了超对称的概念，并沿着不同的路线前进。其中之一是把超对称用于大统一理论，此类理论就称作超对称大统一理论。[13]另一条路线是致力于引力，即所谓超引力理论，它又有不同的形式，有共同之处又有不同的精细结构。所有超引力模型都有一个要点，就是给出了真实世界所可能有的粒子种数，有多少种轻子、多少种光微子、多少种夸克等等，而不是像老的大统一理论里那样没完没了的增多。还没有人能成功地使任何一种超引力理论里的特定种数与真实世界的粒子相符，但比起以前的那种要对付无限多种潜在粒子的窘境来，这看来只是一个相对次要的问题。这些理论中现在较受偏爱的一个叫做“N＝8”超对称，它的拥护者声称它能解释一切，能把力、物质粒子和时空几何装进一个口袋里。但 N＝8 超引力的最好之处在于，它看来不只是可重正化，而且在一定意义上是自我

重正化，那折磨了场论半个世纪之久的无穷大全都在 N＝8 理论中自己抵消，不需要任何人用一个指头去推动。N＝8 对物理学家所问的问题总是给出有限大的答案。真是“超力”！

有点讽刺意味的是，20 世纪 80 年代在寻找可能的方式以把引力和时空装回到粒子物理圈子方面的这些进展，使物理学家想起在 20 年代就曾有过用弯曲时空来解释所有自然力的尝试，在爱因斯坦理论里引力就是这样解释的。这条大统一路线在今后的年代里也许会也许不会被证明为是最好的，但在一本关于宇宙学的书里对它用上少量篇幅是恰当的，因为它最紧密地追随着爱因斯坦在 60 多年以前作出的倡导。

11 维的存在

早在 1919 年，德国柯尼斯堡大学的一位年轻学者特奥多尔·卡鲁查(Theodor Kaluza)，正坐在他书房的桌前研究广义相对论的含义，这个新理论是爱因斯坦在 4 年前提出的，并且已由爱丁顿的远征以一种壮观的方式予以初步证实。他 9 岁的儿子小特奥多尔，在与平常一样静静地坐在书房地板上做游戏。突然，老卡鲁查停止了工作。他又坐了几秒钟，盯着那些写满了方程式的纸，然后轻轻地吹着口哨，双手在书桌上猛拍了一下，站起身来。他再次凝视了一会儿桌上的纸，就开始哼起《费加罗》里他最喜爱的一段，并且一边哼着一边在房间里来回走动。

在小特奥多尔眼里，父亲通常可不是这样的，这个印象深深地留在男孩的脑子里，以至于 66 年后他还能在接受英国广播公司“视界”(Horizon)节目采访时清楚地讲述出来。[14]使他父亲举止反常的那个发现，在几十年无人理睬之后，现在处于对宇宙本质研究的核心地

位。在推敲把引力解释为 4 维时空连续体弯曲的爱因斯坦方程时，卡鲁查作为一位数学家想要知道的是，如果把方程对 5 维情况写出来会是怎样的。他发现，这种 5 维形式的广义相对论也像以前一样包括引力，但还包括另一组场方程，描述着另一种力。那个在小特奥多尔脑子里深深留下烙印的时刻，就是老特奥多尔写出那一组方程并且看出它们很眼熟的时刻，原来那就是麦克斯韦的电磁场方程组。

卡鲁查把引力和电磁力统一到了一起，代价是给宇宙加上了第五维。电磁力看来简单地就是在第五维上起作用的引力。

爱因斯坦"找到"4 个维度(空间 3 维和时间 1 维)来放进广义相对论里时是没有问题的，但遗憾的是并没有证据显示宇宙真的还有第五维。即便如此，卡鲁查的发现还是很惊人并且看来很重要。那时候年轻研究者可不容易一下子就发表惊人的新发现。今天，如果你有一个好的想法，你就写一篇文章寄给一家知名刊物。刊物编辑就把文章寄给一位或几位专家作评价，然后决定是否发表。但那时候认为作者的正确做法是首先把文章寄给一位著名权威，权威如果认可，就把文章加上推荐发表的信一起寄给一个知名的学术机构。于是卡鲁查就把他的结果寄给了爱因斯坦。

爱因斯坦开始时很有兴趣，也很热情。他在 1919 年 4 月写信告诉卡鲁查，他从来没有想到过这个主意，还说"我一看就极为喜爱你的理论"。[15]但后来他开始挑剔一些细节。他自己是个追求尽善尽美者，也就在一系列的信里力促卡鲁查把这些细节整理好再将文章发表。他们之间的通信联系，还有今天看来有点过分的挑剔，一直延续到 1921 年，爱因斯坦突然改变了主意(没有人能确知是为什么)，寄给卡鲁查一张明信片说他即将推荐文章发表。1921 年时对爱因斯坦的推荐是不会有一家刊物编辑表示异议的，于是文章很快就在当年晚些

时候刊登在柏林科学院文集上，其德文标题是“论物理学中的统一问题”。

该文章所表述的理论的明显欠缺(除了第五维并不存在之外)是它没有考虑量子理论，它像广义相对论一样是一个“经典”理论。即便如此，小特奥多尔还记得他父亲的工作在1922年曾引起巨大的反响，但此后就无声无息了。即使是爱因斯坦，虽然把他的余生都用于寻找统一场论，却似乎从那以后也不再理睬卡鲁查的主意。唯一的例外是瑞典物理学家克莱因，本书前面已谈到他的其他工作*，他在1926年找到了一条把卡鲁查的思想吸收进量子理论的途径。

电子或光子或其他微观客体的行为，在量子物理里是由一组有4个变量的方程来描述的。这些方程的标准形式是奥地利物理学家薛定谔首先给出的，故称为薛定谔方程。克莱因重写了有5个而不是4个变量的薛定谔方程，并证明这种方程的解所表示的是在引力场和电磁场二者影响之下的粒子波运动。所有此类把场用多于4维的几何来表示的理论，现在都称为卡鲁查—克莱因理论。[16]早在1926年，这类理论就已经把引力和电磁场结合进了量子理论之中。

这类理论在紧随克莱因的工作之后的那些年里被忽视的原因之一是它们看来并不现实，因为当时并没有更多的力需要考虑。它们所作的“解答”总是引入更多的维度，给方程增加更多的变量以把所有的新场及其载体的作用包括进去，这些新场和载体则全都被描述为与引力一样的几何效应。电磁波(即光子)是在第五维上的波动；而比如说Z粒子就可以是第六维上的波动，如此等等。场越多，力的载体也越多，所需要的维数就越多。但数目不比4种力统一的标准理论如超引

* 参见第七章的注释13。——译者

力所要求的数目更大。

实际上这二者所要求的数目正好一样。超引力理论中排在前头的是 N＝8 理论。它所描述的是如何经超对称运算而把粒子联系于不同的自旋。自旋是量子化的，取半整数值，范围从＋2 到－2。于是从一个极端到另一个共有 8 步(即 8 种超对称变换)，这也就是该理论名称的由来。但还可以由另外一个方式来看。正如卡鲁查修改了爱因斯坦方程要看看它在 5 维情况是什么样子一样，现代的数学物理学家也修改了超引力要看看在它不同维度下是什么样子。结果是，最简单形式的超引力，也是最漂亮最直截了当的数学描述，包含有 11 个维度，不能更多，也不能更少。在 11 维里独有的理论，所描述的可能正是苦苦求索的超力。有了 11 维可用，8 种超对称变换的所有复杂性都消失了，剩下的只有一种基本对称性，即 N＝1 超对称。那么卡鲁查—克莱因理论需要多少维度来容纳所有已知的自然力和相应的场呢？正好也是 11 维，即 4 个熟悉的时空分量和 7 个附加的维度，不能更多，也不能更少。

这一切近来已使许多物理学家大为兴奋，像萨拉姆这样的权威也把粒子和场的世界的这种几何化称为“惊人的、非凡的思想”。[17] 当然要得出一个此种类型的完善理论还有很长的路要走，但是卡鲁查—克莱因理论与超引力的联合看来是当今理论研究最丰产的领域之一。

我们这些凡夫俗子自然会问为什么不能“看见”那些附加维度。但这对数学家来说不是什么问题。每个附加维度都可能由于什么原因而卷曲到自己上面，变得在我们的三维(或四维)世界不可见了。常用作比拟的是水龙软管。从远处看，软管像一条扭动的线，是个一维物体。但靠近了就能看出它是个空心圆柱，是二维物体。扭动线上的每个“点”其实都是一个围绕那个点的圆环，而圆柱就是所有这些环一

个挨一个串成的。在早期的卡鲁查—克莱因理论里，每个时空点也被看作一个小圆环，一个直径只有 10^{-32} 厘米的圆环，在非上非下亦非侧的方向上延展。现代的版本则稍微复杂一点。“圆环”变成了所谓七维球(更准确地说是稍稍压扁的七维球)，即球的七维对应体。但原理是一样的。据数学家说，七维球是仍能允许宇宙如我们所看到的这么复杂的最简单形式的多维结构。

按照这幅新图像，宇宙是诞生于一种 11 维状态，那里没有力与物质的区分，更不要说不同种力的区分了，而只有单纯的 11 维能量。随着能量的耗散，有些维度自我卷缩，产生出我们所认为的物质即“粒子”的构造，即作为在卷曲的维度里振荡的波，并且产生出作为基本几何弯曲的可见表现的自然力。使七维球裂开而展现出灿烂的十维空间所需的能量比大统一能量更高，实际上也就是需要创世本身的能量。

这就是当前研究领域令人晕眩的最前沿，这些新主意在 20 世纪 80 年代的学术刊物上简直如同一股旋风。[18]目前很时髦的一个变种把“粒子”不看作点而看作在多维时空里“运动”的一维弦，这就是“弦”理论。另一方面，有些理论家(令人吃惊的是有霍金，他是 N=8理论的最重要的崇拜者之一，并说过该理论能解释物理学家试图解释的一切，从而可能标志着物理学的终结)却认为所有卡鲁查—克莱因类型的思想都会导致死胡同。这类思想的现代形式也许是太新了，以至于还看不出它究竟会走向何处。毕竟是用了将近 15 年才把非阿贝尔规范的思想有效地运用于某些基本场的统一。卡鲁查—克莱因理论的复活至多只能从 1978 年算起，所以在 20 世纪 80 年代还是童年。如果它真能结出果实，那么至少也得等到 90 年代。

今天那些理论家之所以喜爱卡鲁查—克莱因形式的超引力和超对

称，并不是因为有任何实验证据表明它正确，而是因为它很美和有着内在的一致性。如爱因斯坦曾就广义相对论说过的，它这么美，因而必定是正确的。卡鲁查自己必定也认识到了这一点，因为他是一位优秀的理论家。他儿子曾讲过他是怎么从书上学游泳的。他熟读了书上的理论并相信那是正确的，然后就领着全家来到附近一个湖边，他跳进去，游出了 50 米又游回来，于是他证明了理论果然管用。我们可没有一个合适的湖来把卡鲁查—克莱因理论扔进去，看它是沉没还是能游过去。像萨拉姆一样，我只能说我愿意它是正确的。

但无论卡鲁查—克莱因理论或超引力或超对称能否被证明正确，电弱理论和大统一理论的成功已足以给出对极早期宇宙的新透视。浏览了当今理论物理的最前沿之后，该是兑现本书开头所作的承诺，即讲述这些新理论如何说明宇宙的诞生和极早期演化的时候了。

注　释：

1．得名于挪威数学家阿贝尔(Niels Henrik Abel)。他生于 1802 年，死于 1829 年，对群论这一数学分支作出了重大贡献。他的过早去世是 19 世纪数学的巨大损失。

2．顺便谈一下另一位独立地沿着相似的路线思考的理论家。肖(Ronald Shaw)是剑桥大学由萨拉姆指导的研究生。肖也得到了一个模型，与杨振宁和米尔斯的很相似，但却是在那两位的论文于 1954 年 10 月在《物理学评论》(第 96 卷 191 页)发表之后，他没有费神去争取自己的版本发表。

3．朝永生于 1906 年，卒于 1979 年。他的研究是独立进行的，并于 1943 年首先发表了自己的结果。费恩曼和施温格尔则在二战刚结束时也独立地作出了对 QED 的贡献，朝永的工作为英语世界所知是在 1947 年。他们三人由三条不同途径得到同一个模型，这本身就有力地表明这个模型反映了自然界的某些基本特征。

4．Vol. 19，p. 1264.

5．不要为杨振宁不在 1979 年的授奖名单里而惋惜。他已经因对粒子物理理论的另一个关键性贡献而成为 1957 年诺贝尔物理学奖的获得者之一，他的那个贡献将在下一章关于早期宇宙的故事里讲到。

6．萨拉姆 1968 年的论文不是发表在重要物理刊物上，而是在更不引人注目的诺贝尔专题讨论会文集里，那种文集不在 SCI 收录范围之内，所以对那篇论文没有可作比较的资料。但由该论文的发表形式看来，它不可能比温伯格在《物理评论快报》上的论文受到更

广泛的注意。

7. 数字来自 Pickering，p. 172。

8. 我正好也是在那年出生。有句谚语说，你看到警察开始显得年轻就知道自己在变老。对我来说，我想起自己的年龄是因为看到许多可能戴上诺贝尔奖桂冠的人都比我年轻。

9. 取材于 Pickering，p. 178。

10. 关于搜寻这些粒子的详情可见萨顿（Christine Sutton）的权威著作《粒子连接》（*The Particle Connection*），该书描述了怎样实现这样高能的碰撞以及对观测结果的解释，很有传奇性。

11. QCD 本身除迄今的成功之外也还有未解决的问题。例如，也许有可能或有必要通过希格斯机制给胶子以质量，这肯定会使任何理论家都感到畏惧，因为每次有 8 个胶子，更何况还要自洽地装进一个包里来操纵。

12. 同样还有 W 微子、Z 微子、胶微子和 s 轻子。你可以对此持保留态度。但是欧洲核子研究中心的物理学家在 1984 年末和 1985 年初却为在质子与反质子碰撞实验中观察到的一些事件而大为兴奋，那些事件看来正好要用胶微子或光微子才能最简洁地解释。但如我在 1985 年 5 月所写的，最初的兴奋高潮之后却是一片寂静。也许原先的解释错了，或者是正在积累证据而不久将向全世界公布详情。我们只有等着瞧。

13. 超对称的提出独立并稍早于大统一理论。

14. "What Einstein Never Knew，" 首播于 1985 年。

15. 由 Abraham Pais 引用于 *Subtle Is the Lord*，p. 330。

16. 事实上，努德斯特伦（Gunnar Nordström）1914 年在现在的赫尔辛基大学工作时曾试图找到引力和电磁力的 5 维统一但失败了，曼德尔（H. Mandel）于 1926 年在显然不知道卡鲁查 1921 年论文的情况下独立地提出了与卡鲁查同样的基本思想。

17. *Horizon*，op. cit.；亦见 Salam & J. Strathdee，*Annals of Physics*，Vol. 141，p. 316。

18. 目前的一个兴奋点是所谓十维理论，其中大多数维度都卷缩了，而粒子则被描述为"弦"，故称"超弦"理论。

第九章 极早期宇宙

在宇宙的极早期，即创世后不到1秒钟，粒子物理学与宇宙学会合了。对我来说，这两个物理学分支的会合是在1983年11月，那时我作为一个观察者参加了由日内瓦附近的CERN(欧洲核子研究中心)和ESO联合举办的第一次学术会议，ESO是欧洲南天天文台的简称，其管理机构在慕尼黑，而望远镜却在南半球。会议地点是在CERN，作报告的人中有许多是粒子物理学界和宇宙学界最著名的权威。会议主题是“宇宙的大尺度结构、宇宙学和基础物理学”。虽然会议的内容广泛，包括星系起源和宇宙的终极命运之类的难题，但在我看来核心议题是物理学的统一理论及其对极早期宇宙的应用。原初核合成理论，作为20世纪60年代和我自己参与天体物理学研究时期的重大成就，所涉及的是 $t=0$ 之后几秒到几分钟时的事件；而新一代学者所关心的已不是质子和中子怎样结合成为氦，而是质子、中子、夸克和轻子怎样首次从高能火球中出现。他们已不是简单地把宇宙膨胀作为一个观测事实来接受，而是至少能概略性地给出一族理

论，以解释膨胀是如何开始，在宇宙存在的最早时刻，比起现在来膨胀又是如何快得不可想象。会议的一位报告人霍金，甚至讲述了宇宙最初是怎么创生出来的。

正是会议上这些令人兴奋的内容使我决定写你现在拿着的这本书。从伽利略用望远镜观测天空和赖特对银河本质的猜想开始，我们已经走过了很长的路。最后这两章在全书中所占的篇幅显得较少，但却代表着300年来理论和观测两方面研究成就的顶峰。牛顿说过："如果说我能看得较远，那是因为我站在巨人们的肩上。"[1] 我们现在能看得远至时间本身的起始，这不是因为我们比先辈更聪明，而是因为我们沿着由世代天文学家和物理学家开辟的路径登上了科学的最高峰。

日内瓦会议使我看到了粒子物理学和宇宙学走到一起的很自然的途径。一个简单的例子是巴黎天体物理研究所奥杜兹(Jean Audouze)的报告。他讲的是对早期核合成的理论认识，那是20世纪60年代的老题目，我以为我是熟悉的。但是他所给出的数据的精确性，宇宙学与粒子物理学之间联系的紧密，以及很自然地使人认可这一切的方式，的确都使我大吃一惊。例如，今日宇宙中的氦含量是取决于中子数与质子数之比。计算中所用的一个标准参数定义为中子数与质子数之比的两倍除以1与该比率之和，即

$$2(n/p)/[1+(n/p)]。$$

这个参数除依赖于火球时期质子与中子通过弱作用相互转化的速率外，还与宇宙膨胀的速度有关。宇宙膨胀得越快，在火球中制造氦的时间就越短。但火球中进行的反应并不是只涉及质子、中子、电子和电中微子，而是还会有其他粒子参与，尤其是除电中微子以外的其他中微子的存在会影响氦生成的速率。奥杜兹很有把握地说，每种中

微子的存在都会增大 n/p 比率，从而使上述参数增大 0.1。在有三种中微子的标准模型中该参数值是 0.25。如果该参数值低到 0.23，那就意味着宇宙中至多只能有两种中微子存在，而如果它高到0.26，那就会有四种中微子，但不会更多。

如我已提到的，实验物理学家已经找到了两种中微子，并相信必定还有第三种即作为 τ 粒子搭档的 τ 中微子存在。奥杜兹认为（与会的知名宇宙学家们并未表示不同意见），对宇宙膨胀速率和恒星、星系中氦含量的观测已经好到足以得出结论，即不可能有四种以上的中微子，而最可能的是就只有已经证认出的这三种。

对温度低于 10^{11} K 和时间晚于创世之后 1 秒钟的情况所作的这种相对简单的计算表明，的确只有三对轻子，如果夸克与轻子之间保持着对称性的话，那么也就只有三对夸克可供物理学家研究了。1984 年 7 月，曾发现了 W 和 Z 粒子的那个 CERN 小组报告找到了第六种夸克，即“顶”夸克存在的证据。现在正在进行更多的实验以证实这一探测。*如果宇宙学的确能作出指引，如果观测和实验能进一步互相验证，那么很有可能所有构成物质的终极基元现在都已经知道。正是以这样的成功为基础，物理学家现在已能对质子和中子亦即重子的来由作出解释。

物质的由来

现在大约是把事情放回到用大爆炸的时间尺度来细看的时候了。对于足够小的时间间隔，时间本身的意义变得很不确定，正如位置和动量的意义在量子尺度上变得不确定一样。普朗克曾在 1913 年指

---

* 顶夸克的存在已于 1995 年得到证实。——译者

出，基本物理常数 G，c 和 h 可以经过不同的组合而得出一些数，分别表示有意义的最小长度、最小质量和最短时间。普朗克长度是 2×10^{-33}厘米，普朗克质量是 2×10^{-5}克，而普朗克时间是 10^{-43}秒。一个有着普朗克质量和尺度的“粒子”就是一个量子黑洞。普朗克时间就是光经过普朗克长度所需的时间，约是光“穿过”质子所需时间的10^{-20}倍。

普朗克时间定义了引力的量子效应支配时空结构的时间尺度和距离尺度。普朗克距离尺度是引力变得与其他自然力相等的距离尺度，对应的温度和能量分别在 10^{32}K 以上和 10^{19}GeV 以上。

所以在时间概念本身还能有意义的限度内，宇宙存在的最初 10^{-43}秒的历史是由量子引力支配的。由于尚未建立起令人满意的量子引力理论，目前对那个创世瞬间还无从知晓；由于时间概念本身对那么短的间隔可能已经没有意义，那个瞬间可能会总是无从知晓，它标志着对大爆炸和后来宇宙演化的描述的起点。如果哈勃的红移观测和爱因斯坦的广义相对论的内涵意义真能回推到那么远，那么量子规则仿佛是在告诉我们，宇宙是在“年龄”为 10^{-43}秒时“创生”的。这个创世瞬间是下一章的主题。

创世瞬间后的下一个里程碑是在宇宙年龄约为 10^{-36}秒时。那时，除引力外的其余 3 种力的大统一在温度为 10^{28}K 和能量为 10^{15} GeV 的状态中结束了，从 10^{-43}秒到 10^{-35}秒的这段时间间隔因而就称为“大统一时代”（GUT 时代）。在 10^{-12}秒时（10^{16}K；10^{3}GeV），电弱统一破裂了。那时整个宇宙的温度相当于在欧洲核子研究中心的质子—反质子对撞机里所曾达到过的最高能量。在大约 10^{-5}秒时夸克全都被关进了重子里，那时宇宙的温度已降到 10^{13}K（约 1GeV）以下。从那时往后的历史就由第六章中的标准模型来描述了。但在一个最初

由能量来充满的宇宙里，粒子是只能成对地产生的，即一个物质粒子(如一个夸克)总有一个对应的反粒子伴随，那么又怎会有剩余的物质来形成中子和质子呢？为什么在宇宙已冷却到不能再产生粒子对时，并不是所有的夸克—反夸克对(和电子—正电子对)都湮灭了呢？

解答是在1956年由杨振宁和李政道的工作开始的。李政道也是一位美籍华人物理学家，1926年出生于上海。当时李政道在哥伦比亚大学，杨振宁在普林斯顿，几年前他们曾在伯克利遇见过，并一直保持联系以讨论粒子物理学问题。那时粒子“动物园”正在迅速地增添新成员，有些新发现的粒子的行为很令人困惑。有一族粒子——即K介子的行为似乎不当，它们衰变成稳定粒子(如电子和中微子)的过程似乎违反了粒子物理学的某些规则。李政道和杨振宁找到了一个解决办法，即提出某些弱相互作用可以偶尔违反所谓的宇称守恒的规则。这条规则指的是，自然定律在镜子中反射时是不变的，自然界不区分左与右。但这只是一个假定，从来没有真正检验过。李政道和杨振宁指出了这一点并认为K介子的衰变违反宇称守恒(简记为P)。这使物理学界大吃一惊。只过了几个月，另一位于1912年也出生在上海的美籍华人物理学家吴健雄就用实验对此作出证实，于是李政道和杨振宁共获了1957年的诺贝尔奖，离他们公开发表上述想法只有一年。授奖之快反映了物理学的某个基本思想被推翻所造成的冲击之大。物理学家去除了一个障眼的老教条，开始重新审视粒子世界。

1964年，两位在普林斯顿工作的美国物理学家克罗宁(James Cronin)和菲奇(Val Fitch)把上述结果推广到包括了另一种形式的不守恒。他们用实验证明，K介子的衰变偶尔会允许粒子变成反粒子。这个过程破坏了称为电荷共轭(简称C)的守恒“定律”(其实只是一种基于“常识”的成见)。克罗宁和菲奇还得出，C有时是本身被违

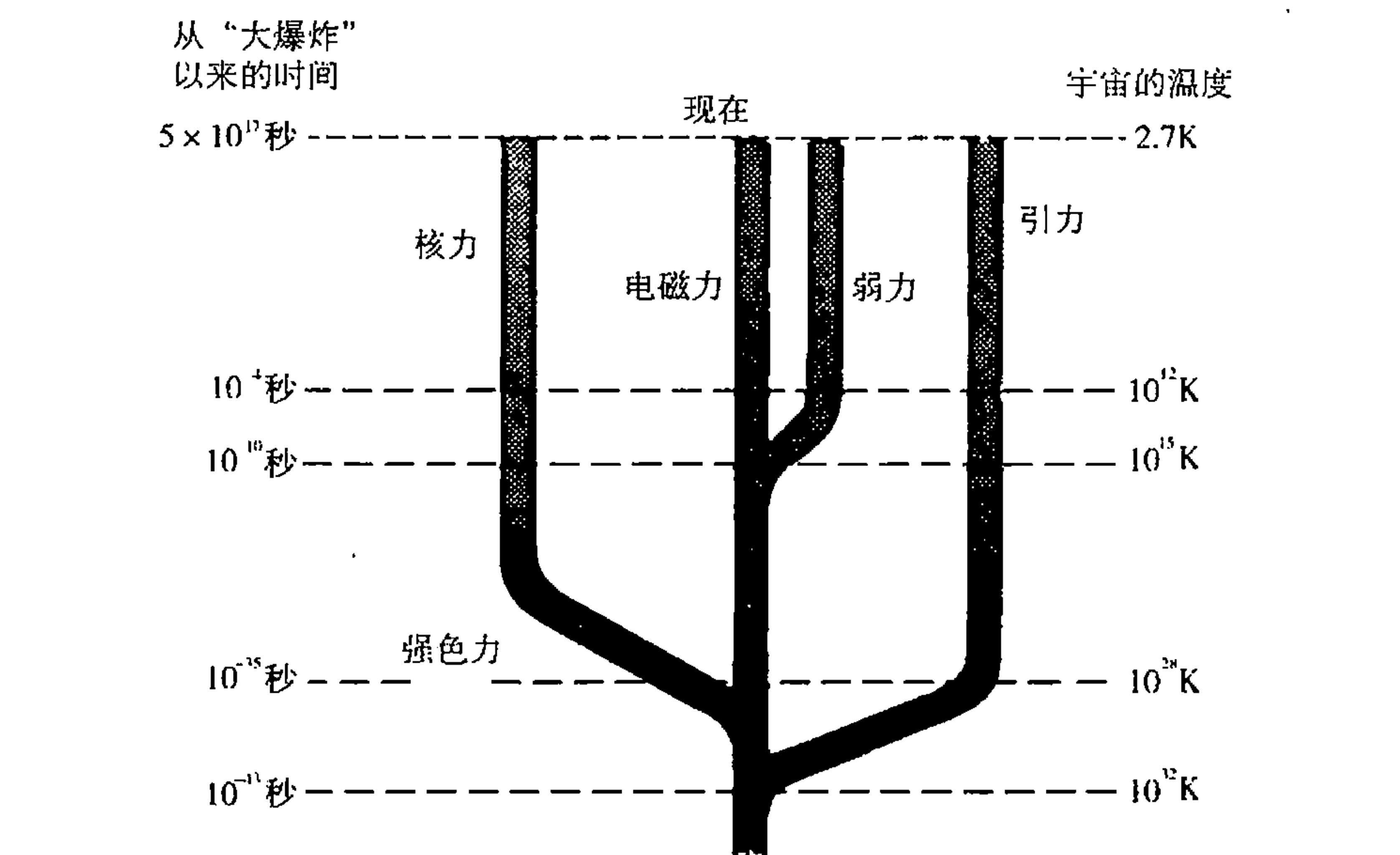


图 9.1 对称性随宇宙的年龄增长和冷却而破缺的示意图，某种原始的超力产生出今日所知的 4 种自然力。

反，而在某些相互作用中，C 和 P 的结合即 CP 也会被破坏。C 的改变并不总能抵消 P 的改变。他们两人也获得了 1980 年的诺贝尔奖。[2]但这给物理学留下了一件有趣的怪事。除了在镜中反射和把粒子与反粒子交换外，唯一还能把粒子相互作用"反转"的途径就是使时间倒流，比如说把一个相互作用拍成电影，然后倒过来放映。把时间(T)的不变性与 C 和 P 的不变性加在一起，就成为一条被称为 CPT 对称的规则。物理学家在 20 世纪 50 年代初就知道，此后的许多实验也已证实，如果对粒子相互作用过程作相当于拍电影的记录，把所有粒子都换成反粒子，把左与右对换，再把电影倒过来放，那么最后结果看起来将与最初的完全一样。物理定律对一个完整的 CPT 变换是不变的，假定 CPT 的每个组成部分在适当意义上独自"对称"之所以曾经显得很自然，原因也就在这里。但事实上，这三个部分的每一个，即 C、P 或 T，都能被单独"破缺"，而合成的 CPT 对称却在任何情况下都仍然保持。例如，由 C 破缺所造成的改变，可以由同一相互作用中

包含了P破缺的改变来抵消；但如果CP不变性被打破了，唯一能保持整个CPT系统对称性的途径就是T也破缺，并且正好与CP破缺相抵消。

对粒子来说，这是什么意思呢？这可以再次由K介子尤其是其中的一种即K_L^0看出来。这是一种中性粒子，像光子一样，它也是自己的反粒子。它能以两种重要方式衰变。一种衰变产生一个负π介子、一个正电子和一个中微子。另一种则生成一个正π介子、一个电子和一个反中微子。别担心π介子，它们是玻色子，不受那些在绝大多数情况下对费米子似乎都适用的守恒定律支配。但是，那些守恒定律真在任何时候都对费米子适用吗？如果物理定律是对称的，那么在大量K介子中就应该是两种衰变方式各占一半，从而保持宇宙中电子与正电子之间的平衡。但事实上第一种衰变稍稍多一点，也就是说K介子衰变会极少量地增加宇宙中的正电子成分。

这是该水平上物理定律对称性的第一个缺口。宇宙似乎表现出稍稍偏爱用左手，至少在事关电子时是如此。那么对重子和强作用来说又怎样呢？在20世纪60年代中期还没有一种好的强力理论，所以对质子和中子的计算也就无一可与李政道和杨振宁所作的计算相比。夸克概念在茨威格和盖尔曼脑中还只是隐约闪现，当时没有、现在也看不出有什么办法可望能由地球上的实验来直接检验强作用理论。在这样一种似乎很难有所作为的背景下，杰出的苏联物理学家萨哈罗夫(Andrei Sakharov)却于1967年定出了一套基本原理，它们适用于在早期宇宙中优先产生正物质粒子的任何过程。[3] 萨哈罗夫说，要使大爆炸中产生的正物质多于反物质，有三个条件必须满足。第一，必须有非重子生成重子的过程。第二，这些重子反应，至少是其中要紧的那些，必须违反C和CP守恒。否则即使重子由某种过程产生了，也

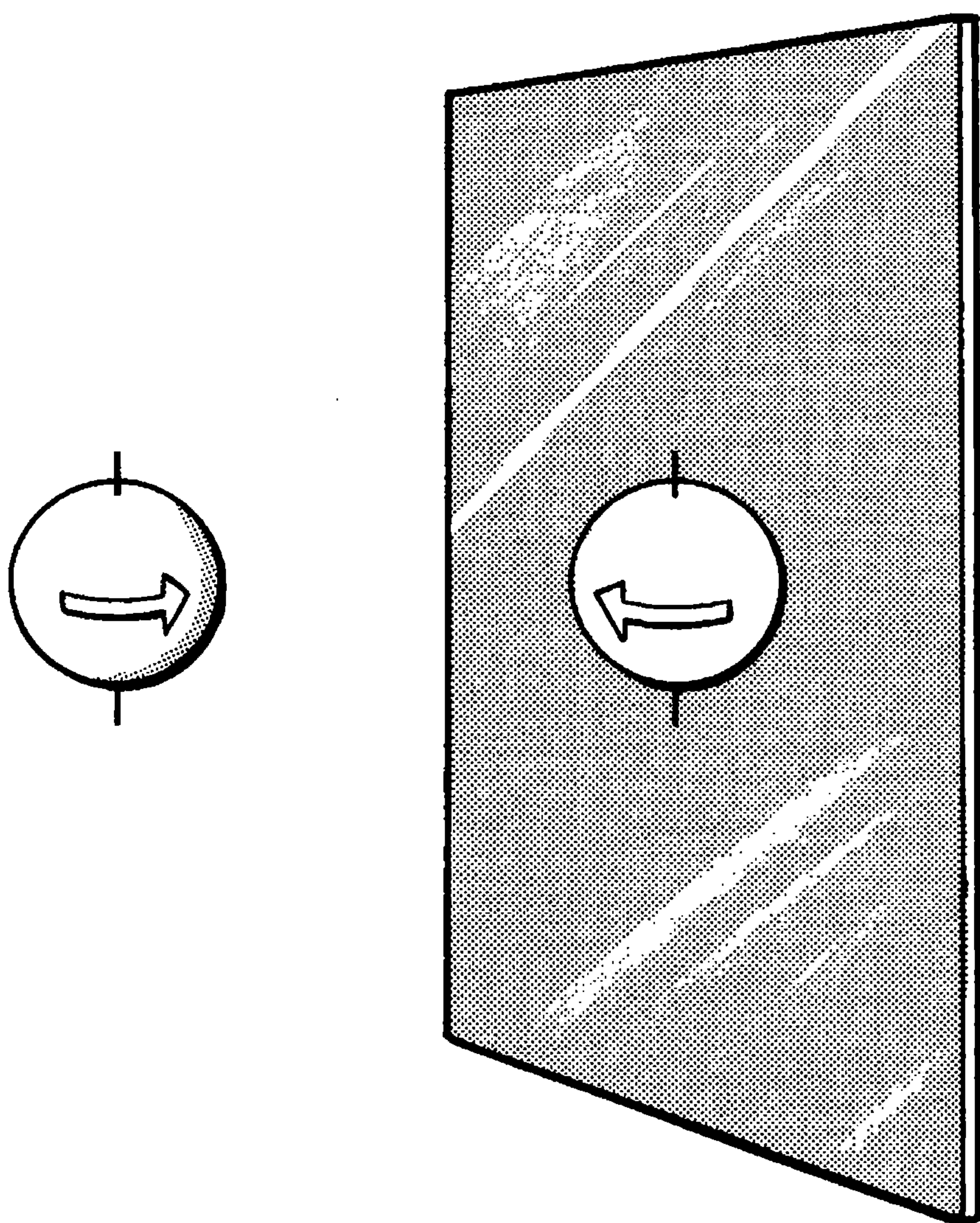


图 9.2　在镜中世界里，球原来的运动可以由时间倒流来恢复。镜反射和时间流两者都是不对称的，但它们合在一起能产生对称性。CPT 对称也是这样由无需单独对称的部分组合而成。

会有等效的过程产生相同数量的反重子，结果粒子与反粒子还是会相遇并湮灭。第三，宇宙必须由热平衡态演化成非平衡态，也就是说必须有一个确定的时间流使 CP 过程不守恒，而 CPT 保持守恒。

对宇宙中“时间之箭”的这种需要从物理上是容易理解的。如果辐射与物质在某种均匀的高温下处于平衡，那么所有把辐射变成粒子

和把粒子变成辐射的过程，或者把粒子变成别的粒子的过程，也都处于平衡，在每个方向上都能同样顺畅地进行。要使今日宇宙中有正物质留下来，最终以能使正物质有剩余的那种方式来衰变的粒子就必须在宇宙的这种高温状态中产生出来。但只要温度仍然极高，则生成正物质的反应就总是被变回到原初粒子的逆反应所平衡。只有宇宙温度降低，反应的平衡才会朝有利于低能量状态的方向倾斜，也就是朝有利于形成剩余正物质的衰变方向倾斜，而宇宙由高温态到较低温态的演化也就给出了时间的方向。所以，要得到宇宙中的物质，就需要一个高能态作为开始，再加上向低能态的移动。

萨哈罗夫是只以基本物理原理为基础来陈述这些条件的，他并不知道物质的创造过程包含什么力和粒子，也不知道究竟在什么温度下反应之平衡才会倾向于对物质有利。他的思想远远超越了时代，文章又是用俄文发表的，因而在 20 世纪 60 年代后期并没有产生重大影响。这些思想在静待 10 年之后于 1978 年被日本东北大学的物理学家吉村(Motohiko Yoshimura)重新提出。吉村一直在研究规范理论和大统一理论(GUT)。所有大统一理论的唯一最确实的预言是有 X 玻色子，其质量约为 10^{15}GeV，即质子质量的 10^{15} 倍。由于无望在地球上的加速器得到这样大质量的粒子，吉村提出 X 玻色子的存在能以其在 GUT 时代对宇宙的影响揭示出来，那个时代是在 10^{-35} 秒之前，X 玻色子是物质的主宰成分。遵照着吉村的指引，后来又知道了萨哈罗夫的开创性工作并给了他应有的荣誉，理论家们很快发现，X 玻色子在 GUT 时代结束时能造成重子相对于反重子的剩余，正如 K 介子衰变现在能造成正电子相对于电子的剩余一样。这种剩余很小，但已足以胜任。

前面已讲过 X 玻色子在今天如何使质子衰变。任何能消灭质子

的过程都有造出质子的反过程与之对应，这正是审看 GUT 时代 X 玻色子反应的一条途径。当然，所有这些反应真正涉及的是夸克和轻子，而不是整个质子，但原理是一样的。一个上夸克能发射出一个 X 玻色子而变成反上夸克，X 玻色子又衰变成一个反下夸克和一个正电子。GUT 给定了把夸克变成轻子以及由 X 玻色子生成夸克和轻子的处方，萨哈罗夫三个要求的第一个满足了。

在 GUT 时代，X 玻色子本身也与充满宇宙的辐射处于热平衡。X 和反 X 对被不断地造出来，又不断地相遇而湮灭。但到了 $t=0$ 之后的 10^{-36} 秒或最晚是 10^{-35} 秒时，被产生出来的新对子数就不再能够与已有的 X 和反 X 粒子的消耗速率相配，因为宇宙已冷却到 10^{15} GeV 以下。假如消耗过程只是对湮灭，那么故事也就完结了，宇宙也就会只是除了冷却着的电磁辐射外一无所有的一个膨胀时空。但是按照

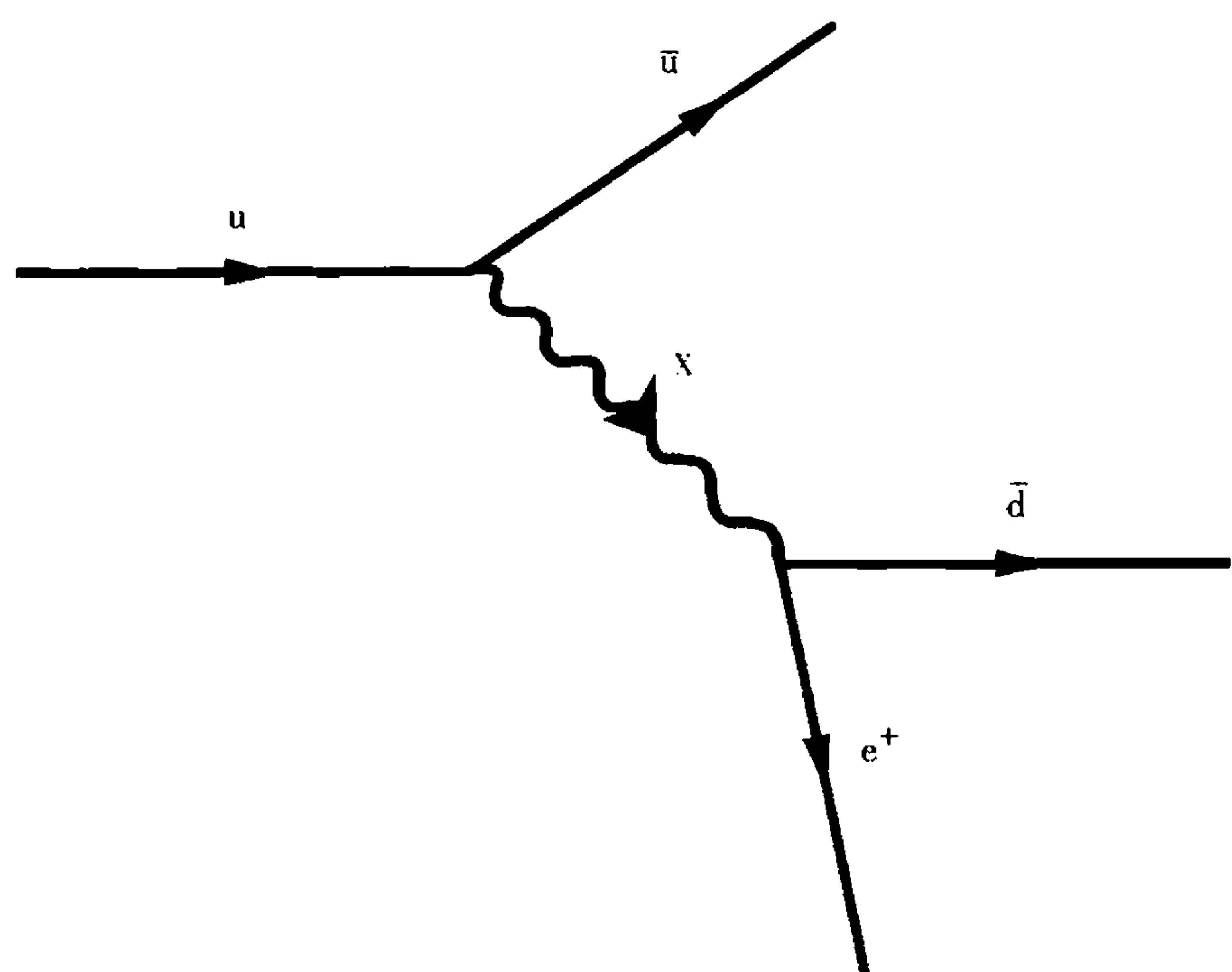


图 9.3　包含了 X 玻色子的反应能由非重子生成重子，也能反过来。在可以得到足以使 X 玻色子存在的能量时，夸克与轻子能相互转化。

GUT，X 粒子也能衰变成夸克和轻子，而且它们中有许多在湮灭之前就这样做了。

X 玻色子衰变成其他粒子有两种主要方式。一种是生成两个夸克，另一种是生成一个反夸克和一个轻子(实际上由于其巨大的质量能量，一个 X 玻色子就能衰变成一大批夸克对和夸克—轻子对，这里还是一次只讲一步)。反 X 的衰变也就相应地生成反夸克对或由夸克和反轻子组成的对。但由于衰变过程不遵守 C 和 CP 对称，反 X 和 X 的衰变速率是不一样的，于是萨哈罗夫的第二个要求又满足了。很粗略地讲，每当反 X 衰变生成 10^9 个反夸克时，相应的 X 衰变会产生 10^9 加 1 个夸克。这就是今日宇宙中所有物质的由来。随着宇宙的膨胀和冷却(符合萨哈罗夫的第三个要求)，每 10^9 个夸克中就有 1 个在 GUT 时代后找不到反夸克伙伴，因此就再也不会湮灭，而是留下来形成作为今日宇宙中物质主体的质子和中子。[4] 我们大家能够存在，是多亏了物理定律极其微小的不对称，多亏了 X 玻色子衰变时物质相对于反物质的优势，这种优势至多只是对每 10^9 个反夸克多出了 1 个夸克，也就是在 GUT 时代以 X 和反 X 对的形式存在的所有物质中的 $1/10^9$ 的不平衡。

10^9 这个数不是信手拈来的。它是由物质密度和宇宙背景辐射强度计算出来的今日宇宙中光子数与重子数之比。还没有一个大统一理论好到能精确预言在 GUT 时代末尾物质与反物质之间的不平衡“应该”是多大，对光子数与重子数比率的估计只能给出一个相当大的范围，即可能是 10^4 到 10^{13} 之间的任何值。至少所“要求”的值是在这个范围的中央。最重要的是，GUT 原则上给出了造就宇宙中物质的机制。这正是宇宙学家在 20 世纪 70 年代末和 80 年代初对新的高能物理学表现出巨大热情的原因之一。他们想，也许高能物理还能解开

有关标准宇宙学模型的其他一些深刻奥秘。他们很快就被证明是对的。

大爆炸的问题

宇宙学家对GUT能解释物质起源这个消息的热烈欢迎，部分地还因为他们由此看到了解决一些存在已久的问题的可能性，那些问题曾被放在一边，只是因为不知道对它们如何下手。标准大爆炸模型本身无疑是成功的，比起能洞察宇宙的起源这一惊人成就来，它所带来的问题显得是次要的。但是随着时间的流逝，那些问题对宇宙学家的折磨非但没有减轻，而且到了20世纪70年代末它们看来比1965年时还要严重得多。如科学史上常见的那样，一代研究者在取得重大进展的同时也遇到一些新的困难，这些困难就留给下一代人去冲击和解决。

除了现在看来已经解决的物质起源问题外，标准模型还有几个重要的缺陷。第一，有一个被称作"视界"问题的困难。它的出现是因为宇宙在所有方向上看来都是一样的。由星系和星系团在天空中的分布看来，宇宙是很均匀的；但真正体现宇宙均匀性的是背景辐射，它是各向同性的(即在各个方向上都一样)，偏差不超过万分之一。来自天空中一处的辐射怎么"知道"必须有多强，才能与来自天空中相反位置的辐射(其实也与来自两者之间任何一点的辐射)如此精确地相等呢？我们今天所作的观测是我们与来自宇宙的相反"两侧"的辐射的第一次接触，而按照标准模型，那些区域从来没有相互联系过，因为它们总是相隔得比光在宇宙年龄这么长的时间内所能走过的距离还要远，在宇宙迄今为止的每个历史时期都是如此。如果如标准模型所给出的，3K辐射是$t=0$之后300 000年时发出，那么从地球上看来，天

空中就只有张角不到2度的区域在当时有相互联系。所以背景辐射就“应该”是拼凑成的，有着尺度为2度左右的颗粒状结构。看来宇宙是以一种完全平整的状态从火球中诞生出来的，到处都“内定”了精确相等的能量密度(即同样的温度)，即使那些相隔太远、连以光速传播的信号都来不及穿越其间的区域也是如此。但大爆炸中温度的这种均匀性又是怎么来的呢?

视界问题直接引起了第二个问题，即星系的存在。宇宙既是诞生于如此平整的状态，星系这么大的团块又怎么能形成呢?在一个均匀地膨胀着的完全平整的宇宙里，每个物质粒子都会离其余的每个粒子越来越远，根本就没有能在引力拉拢下长成更大物质团块的种子。并不需要很大的种子，在 $t=0$ 后500 000年时，只要有那么点不规则性使某些区域的密度比平均值高出0.01%就够了。但即使这么小的不均匀性又怎么能从完全平整的大爆炸中形成并显现出来呢?

接下来是磁单极问题。成功地预言了宇宙中物质数量的大统一理论，还预言以有质量也很容易探测的粒子形式存在的磁单极。应该能在宇宙线中找到它们，但没有找到过。迄今的实验中只有两次事件有可能是由磁单极在通过实验装置造成的，但在没有更多的证据时不会有人把这作为肯定的结果来接受。

但是，激发了新的研究浪潮、并导致宇宙在 10^{-30} 秒之前演化的某种新理论的问题，是所谓“平坦”问题。这要回到对星系、红移和宇宙膨胀的老式研究。事情是这样的，我们可以算出宇宙膨胀的速率，还可以通过对星系计数来估计宇宙中物质的量，或者物质的密度，密度才是重要的。爱因斯坦方程允许的可能性是，宇宙开放，从而注定永远膨胀；或者闭合，从而终将收缩回到火球。或者，仅作为一种可能性，宇宙是平坦的，正好在前两种可能性之间的引力刀锋

上，如在第四章中已讲过的。

观测所显示的宇宙实际密度，肯定是在使宇宙闭合所需要的临界密度的0.02倍到10倍的范围内，宇宙学家对实际密度究竟是在临界值的哪一侧已争论了多年。但他们在争论过程中慢慢认识到，有争论的余地这个事实本身就是宇宙最值得注意的特征之一。为什么密度观测值不是使宇宙闭合所需临界值的 10^{-4} 倍或 10^{6} 倍或其他与1相差很大的倍数，从而由观测就能明显地作出结论呢？今日宇宙实际上是非常接近于似乎最不可能的状态，即绝对平坦。而且这意味着它必定是诞生于一种甚至更为平坦的状态，这一点是两位与发现3K背景辐射有关的天文学家，即普林斯顿的迪克和皮布尔斯于1979年指出的。

发现今日宇宙是处在一种哪怕是近似的平坦状态的可能性，甚至比发现一枝铅笔笔尖朝下倒立几百万年不动的可能性还要小；如迪克和皮布尔斯所指出的，宇宙在大爆炸时从平坦状态的任何偏离都会随着宇宙的膨胀和年龄增长而显著地变大又变大。就像平衡在笔尖上的铅笔受到极小触动时会发生的一样，宇宙也会很快离开完全平坦的状态。完美的均衡态是一种平衡，但却是不稳定的平衡，任何细小的偏离都会对其带来灾难。我们可以在想象中把钟倒拨，计算出宇宙在火球时期必须有多平坦，才能在今天仍有如此接近临界值的密度。迪克和皮布尔斯，后来还有其他人，已经为我们作了这种计算。如果宇宙现在的密度是恰使它闭合所需要的值的十分之一(大多数天文学家会同意这是依据可见星系所作的合理估计值)，那么在创世后1秒时宇宙的密度与当时的临界值在 $1/10^{15}$ 的精度内相等。如果退回到GUT时代结束即在 10^{14} GeV和 10^{-35} 秒时，那时的密度必须只比临界值小 $1/10^{49}$。这不可能是一个巧合，而是必定意味着物理定律要求宇宙从大爆炸中诞生时处于某种极其平坦的状态。

1979 年春，迪克在康奈尔大学作了一个报告，讨论了平坦问题，并指出宇宙是何等接近于不断膨胀与猛烈再收缩之间的平衡态。听众中有一位年轻的研究者，名叫古思(Alan Guth)，是个不怎么情愿的宇宙学家，被他的朋友也是在康奈尔的同事泰伊(Henry Tye) 拖来研究有关磁单极的一些问题。古思后来回忆说，他开始搞宇宙学完全是为泰伊所迫，泰伊“不得不施加许多压力，因为那时我很坚决地相信，宇宙学是这样一种领域，任何人想说什么都行，没有人能证明他错了”。[5] 尽管开始时是抱着如此怀疑的态度，古思后来却发觉宇宙学的那些思想很有吸引力，于是逐渐深入到研究之中。古思用了几个月时间把迪克在报告里讲述的思想与自己从宇宙学和粒子物理学(那是他原来的专业)里得来的其他思想结合起来，到了那年 12 月他开始有了明确的主意。12 月 6 日的下午、晚上并直到深夜，古思把他的想法整理成了自伽莫夫 30 多年前的工作以来对早期宇宙学的第一个重大的理论新贡献。他把一个全新的概念引入了宇宙学，这一步的意义如同宇宙蛋的思想一样深刻。

最后的免费午餐

古思是沿着常规途径在 20 世纪 70 年代进入物理学研究的。他于 1947 年出生在新泽西州的新不伦瑞克，3 岁时全家移居到新泽西州海兰帕克，在那里上完高中，1964 年进了麻省理工学院学物理。他完成了学士和硕士学业，1972 年又得了博士学位，然后就碰上在物理学界找一个稳定研究位置的问题。我可是太知道 20 世纪 70 年代初的那种问题了，因为我是在 1971 年完成博士学业的。那时有大量才拿到博士学位不久的学物理的年轻人，其中大多数比我强得多，而稳定的工作位置却很少。唯一的办法是一边在一个短期位置上干着一边找下一

个，盼望着有一天能得到长久职位。我没有这么苦干，而是不做研究专事写作。古思显然比我更能坚持(也更有才干)，他在普林斯顿大学和哥伦比亚大学干过之后又到了康奈尔大学，就在那里听了迪克 1979 年春天关于平坦问题的那个报告。1979 年 10 月他又离开了康奈尔，去斯坦福直线加速器中心干了将近一年。就在那里由迪克的报告播下的种子开始生长了，古思自己在那些日子里收集的宇宙学和粒子物理学资料则提供了所需的养分。

古思清楚地记得那一天，汇集在他脑子里的各种思想突然一下子结合起来了。那是 1979 年 12 月 6 日星期四，突破是由下午与一位来自哈佛的访问者科尔曼(Sidney Coleman)的谈话开始的。他俩讨论的是 GUT 的最新进展，有物质由 X 玻色子产生出来的问题，还有许多其他问题。到了晚上，古思觉得脑子里有什么东西要结晶出来。在家里吃了晚饭后，他坐下来试图把在脑子里奔涌的想法用数学式子写在笔记本上。妻子早已入睡了，他一直干到深夜，写下了一页又一页整洁而细致的计算。在我写这本书时，古思给了我那一本和后来一本笔记上几页的照片。在笔记本上从 12 月 6 日到 12 月 7 日凌晨所写部分的末尾，有一页的上方是一个仅 5 行的小段落，周围仔细地用双线画上了方框，标题是大写的**“惊人的认识”**，内容是：“这种超冷却能解释为什么今日宇宙平坦得难以置信，因而解决了迪克在爱因斯坦纪念日报告中所指出的微调佯谬。”古思所发现的是对极早期宇宙描述的基础，即所谓暴胀，这个名称是他自己不久后取的。几星期后，这个在宇宙学领域显然还称不上是专家的古思，在同加速器中心一位物理学家温斯坦(Marvin Weinstein)的谈话中首次得知，宇宙学家还在为宇宙大尺度的均匀性即视界问题所困扰，他意识到自己的新模型也能对此作出解释。他虽然在后来的研究中发现自己的模型至少在一个方

面有缺陷，但还是写成了论文发表[6]，并且在文中明确表示，希望别人能弥补那个缺陷并把基本思想向前推进。紧随关于暴胀的工作之后，古思当上了麻省理工学院物理系的访问副教授，接着在1981年6月又成了正式副教授，至今仍在该职位上。

为什么这个模型，不论是否有缺陷，会使得古思如此兴奋，并且在20世纪80年代带来了宇宙学的一场革命呢？首先是由于那个在GUT被破坏时的希格斯转移的思想，它现在已为粒子理论家所熟悉，指的是对称性随着中间矢量玻色子“吞食”希格斯粒子并获得质量而破缺的过程。在许多大统一理论里，当希格斯场为零时对称性是不破缺的，但一旦有至少一种希格斯场不为零，对称性就会自动破缺。希格斯场有一个特性，以前已经讲过。场的最低能量状态并不一定对应于场的值为零。古思把希格斯场为零的状态叫做“假真空”，而用“真真空”来指有着更低能量密度的状态，在这种状态中希格斯能量为最低值，但场本身不为零，对称性被破缺，矢量玻色子获得质量。他主张，GUT破裂对应着从假真空到真真空的转移；当他把一些数字放进方程式中时，使人大吃一惊的结果出现了。

当能量密度很高时(如果我们正在考虑早期宇宙，那么情况就是如此)，最低点的精确值和位置都不很重要。只有在温度下降时场才会安定在某个或别个最低态。还是用光滑大碗里的弹子来作比拟。如果弹子是快速地从碗边滚进碗中的，也就是说有着很高能量，它就能在碗里边转边上升，就像杂技表演的飞车走壁一样。最低点的位置，或碗的深度，对弹子的行为没有影响。但随着能量的减小，弹子在碗里转得慢了并往下落，最后停在碗底，也就是最低能量态。现在最低点的位置就是最重要的了。与此类似的事情在宇宙从10^{-43}秒冷却到10^{-35}秒(那时的宇宙温度约是10^{27}K)的过程中也曾发生。希格斯场也

逐渐地安定在一个最低能量态。但究竟是哪一个？假设那些场是处于一个对应于假真空的状态，那就只是一个局域最小，像火山口的凹坑那样。在这种状态仍有着未破缺的对称性，强力与电弱力仍是等同的，场的能量密度高到每立方厘米 10^{95} 尔格（10^{88} 焦），是原子核能量密度的 10^{59} 倍。

这种状态可以与水的过冷，即冷却到 0℃ 以下而不结冰相比。随着不断地冷却，水最终会相当突然地结成冰，并释放出熔化的潜热。在 0℃ 以下，冰是能量更低因而更稳定的状态，但是伴随着潜热能量释放的向更稳定状态的转变，并不总是发生在 0℃。古思认为类似的事情也可以对希格斯场发生。虽然在宇宙冷却到 10^{27} K 时对称性应该破缺，但场却可以在假真空态停留一段时间，就像过冷水在冰点以下仍保持为液体一样。只有在场能够有一种方式穿过周围的能垒并到达真真空态更深的最低点时，对称性才被破缺。

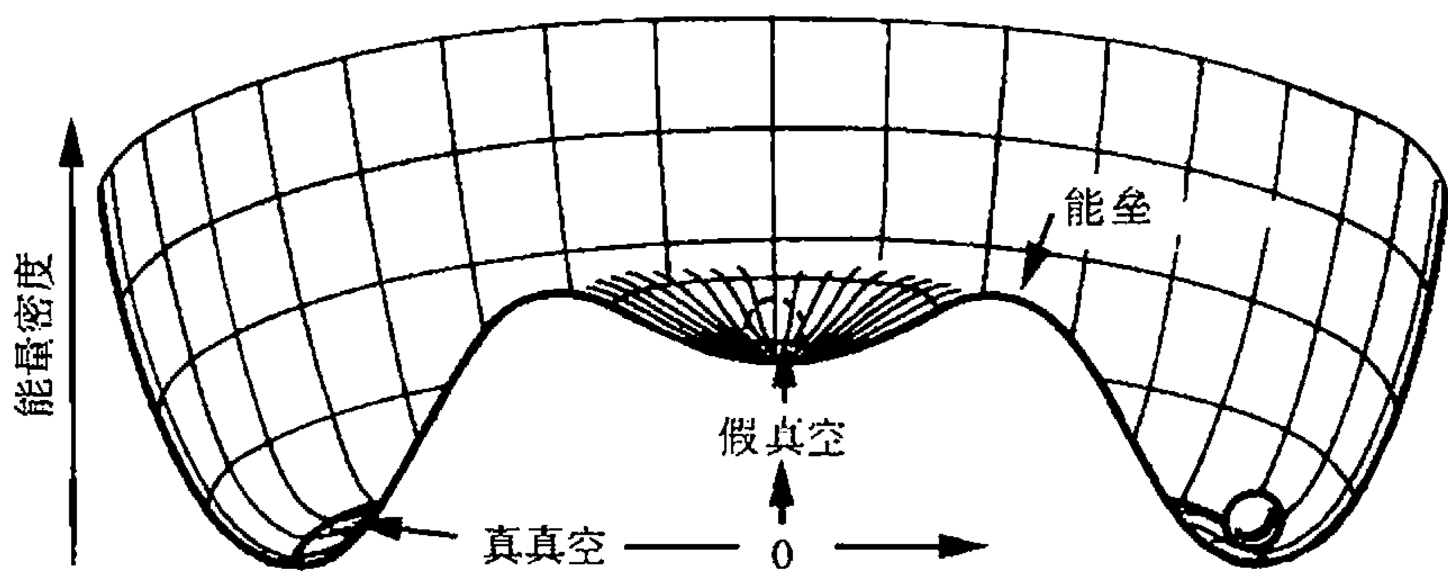


图 9.4　在第一个版本的暴胀模型里，宇宙先被囚禁在希格斯场的“假真空”态，然后像 α 粒子逃出原子核那样，穿过能垒落到真真空态，这个过程中释放的能量驱动宇宙快速“暴胀”。

与火山的比拟当然也是大家熟悉的，我曾用来描述囚禁在原子核里的 α 粒子。正如 α 粒子能借不确定原理之助逃出束缚那样，希格斯场也能靠不确定原理帮忙而从假真空态穿出。但古思惊讶地发现，场被囚禁在假真空态或过冷态时所具有的巨额能量密度，能够给宇宙

一个强大的向外推动，使它膨胀得比标准模型里快得多。这个效应在一段短时间里就好像有一个比爱因斯坦的任何想象都有力得多的宇宙常数。结果是，宇宙作指数型膨胀，每过 10^{-34} 秒其尺度就增大一倍。这听起来好像是个很小的膨胀速率。但每过 10^{-34} 秒就增加一倍，也就是在 10^{-33} 秒里加倍 10 次，尺度就增大到 2^{10} 倍，在 10^{-32} 秒里就增大到 2^{100} 倍。[7] 在比一眨眼短得多的时间里，一个大小只有质子的 10^{-36} 倍的区域能够暴胀到 10 厘米的样子。暴胀把极早期宇宙那个渺茫的亚微观世界很突然地带到了我们所熟悉的尺度。

但是一旦希格斯场到达了真真空态，快速的指数暴胀就停止了。场的能量转为产生出巨大数量的粒子对，并把它们再加热到将近 10^{27} K。实际上我们所知道的宇宙中所有的物质和能量可能都是这样由暴胀产生出来的。这种可能性的出现是因为宇宙中的引力能是负的，而且宇宙越大就越负。古思指出，暴胀过程中能量应该是这样守恒的，即宇宙的物质能量增大，而引力能变得越来越负，两者几乎是精确地抵消，用他的话说这是宇宙“最后的免费午餐”。随着宇宙膨胀减慢到标准大爆炸模型的正常步调，由希格斯场生成的 X 玻色子的衰变就造成物质相对于反物质的微小超出，此后的事情就都如第六章中的标准模型所述了。暴胀在远未到 $t=0$ 之后 10^{-30} 秒就结束了，但它已经解决了视界问题、平坦问题，还有磁单极问题。

视界问题的解答简单地就是，今日宇宙中“相反两侧”的区域在宇宙刚诞生时确实是“相连”的，然后才由于暴胀而分开。我们今天看到的宇宙之所以特别均匀，只因为它是由一粒极小极小的种子变来的，在那种子里所有能量是均匀分布的。对平坦问题的解答稍稍更精巧一点。吹胀一个气球时，气球面的弯曲程度越来越小，越来越接近于平面。暴胀过程中时空曲率的变化也是如此。无论开始时曲率是多

大，到时空已膨胀了10^{50}倍时就与平坦宇宙不可区分了。再弯曲的宇宙，当它膨胀到一粒葡萄那么大时，在我们的任何观测所能揭示的程度上，就已经变成平坦宇宙了，其密度已非常接近于临界值。磁单极问题也能很简单地解决。尽管在暴胀之前的极早期宇宙里可能有许多磁单极，我们现在所能看到的空间区域却是从如此微小的体积膨胀而来的，以至于其中连一个磁单极都没有包含。

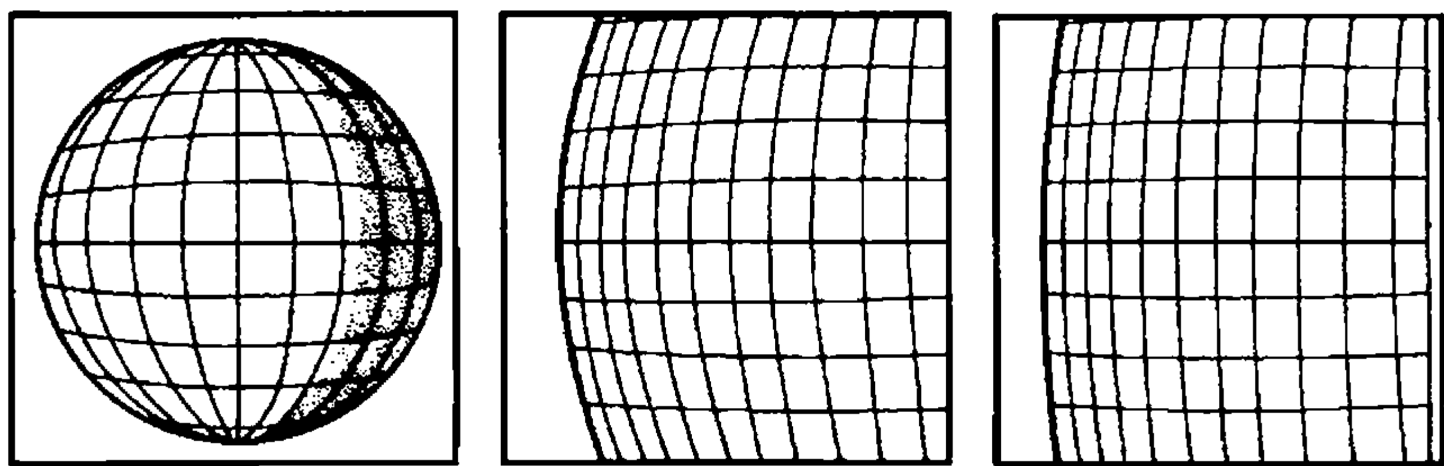


图 9.5　无论开始时曲率是多大，暴胀都使宇宙变得平坦。图中自左至右三个球的半径依次增大 3 倍，故总效果只是“平坦”了 9 倍。而暴胀使宇宙变平坦了10^{50}倍。

对磁单极问题的这个解答显示了古思模型和后来的暴胀模型的一个关键特征。那就是，在我们所看到的宇宙之外可能还有别的时空区域，它们不是在与我们的宇宙“泡”同样的时刻也不是以同样的速率暴胀来的。我们这整个可观测的宇宙可能只是某个大得多的总宇宙的“局部”区域。但是别的那些区域有可能被观测到吗？这可就是最初的暴胀模型的一个严重缺陷了。

如古思所意识到的，整个时空都可能随机地发生从假真空到真真空的转移。于是就可能有许多暴胀来的时空泡，每个都有自己特定的希格斯场值，每个都有着稍稍不同的物理定律，因为在各个泡中对称性的破缺方式稍有不同。在最初的暴胀模型里并没有什么精巧的办法能使暴胀平稳地发生，相反倒是能有许多不同时空区域的暴胀泡以不同方式从假真空中穿透出来。泡也会成团，就像池塘里成团的青蛙

卵，或者像海绵里的孔，而泡之间的边界应该有很强的活动性，从而是一个可探测的特征。但我们的宇宙看来完全不是这样子。所以古思最初的模型的这个预言显然是错了，模型出现了裂缝。但是它的指数膨胀特征是这么好，解决了这么多问题，使得许多宇宙学家都愿意它是正确的，而不管有什么明显缺陷了。对这种解决宇宙学主要难题的撩人而又有缺陷的可能性的苦恼，是由莫斯科列别捷夫研究所的林德(Andrei Linde)为我概括的，正是他在 1981 年取得了下一个重大进展。

林德于 1948 年出生在莫斯科，在莫斯科大学学物理，后来到列别捷夫研究所做研究，开始时是在基尔日尼茨(David Kirzhnitz)的指导之下。林德研究的是包含希格斯场的高能相变的本质，他知道了古思的工作感到很高兴，因为他和基尔日尼茨一起搞的课题看来可能有宇宙学的意义，但他对解释如何能由假真空平稳地转变到真真空的困难颇感懊丧。他在 1985 年 3 月写给我的信中说："一直到 1981 年夏天我都觉得浑身不适，因为我看不出有什么办法能把情况改善，而我又不相信上帝会误失这么一个使创造宇宙的工作简化的好机会。"但就在 1981 年夏天，他想出了对问题的一个解答。他的建议简单说来就是，假真空态并不是一个像火山口那样的深阱，而是一个能量的平缓高原，由中心向外缓慢倾斜，最后落到真真空态。这样一个高原上的希格斯场会很从容很平稳地"滚出"而进入真真空，没有那种与钻过能垒相伴的一系列局域量子跃迁造成的混乱。结果就是一个平滑均匀的宇宙，一个像一团果冻而不是一把青蛙卵的宇宙。这个主意后来被称为"新暴胀方案"，尽管它需要一些时间来理解，但到 20 世纪 80 年代中期已是大受欢迎。

林德首次出示自己的模型是在 1981 年 10 月莫斯科的一次国际讨

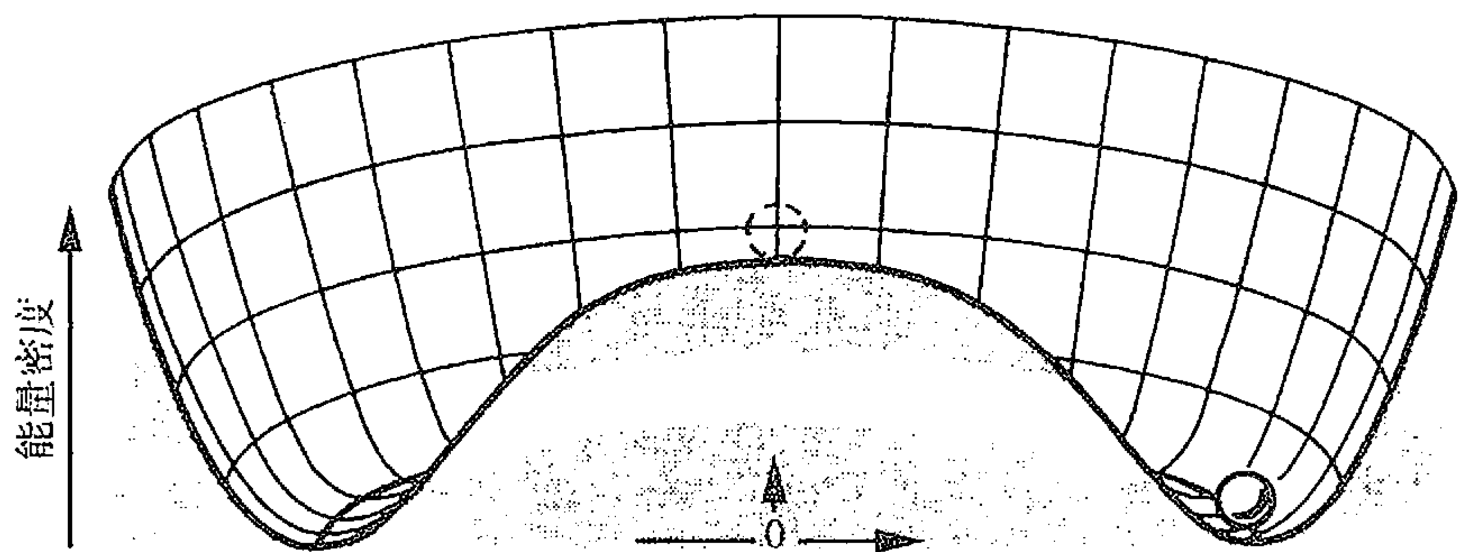


图 9.6　按照第二种暴胀模型，宇宙从能量高原顶部缓慢地滚落到真真空态。有关方程式再次表明，伴随的能量释放能够驱动整个宇宙的短暂而又极快的指数式膨胀。

论班上。当时在场的霍金曾予以“驳斥”。但再次考虑之后他发觉林德的想法是有道理的，于是后来在费城作的一个报告中提到了林德模型，并就此课题与一位同事莫斯(Ian Moss)合写了一篇论文。费城的两位研究者阿尔布雷克特(Andreas Albrecht)和斯坦哈特(Paul Steinhardt)独立地得到了与林德同样的结论，并于 1982 年 4 月发表，他们的论文中引用了林德的独立工作。大致就在这时“新”暴胀假说才得到重视(科学家总喜欢对一件事听说了两次才相信)，而古思主题曲的林德变奏此后几年在众多的暴胀模型中也就居于领先地位。

从 1982 年起有许多人对新暴胀假说做了进一步的修补，试图把它微调得能与宇宙的观测特征相符。有一种修补是这样的，假真空态先由“高原”中心附近的一个小能垒包围着，然后钻出来形成许多泡，这些泡先在高原上各自演化，然后逐个地合并到一起，滚落到真真空态。每个时空泡只要还在高原上就保持加速膨胀，每个泡都能很容易地增长到可观测宇宙的尺度。在这一切发生的时候，也就是在相变开始的时候，光从 $t=0$ 起已能跑过的距离(即“视界距离”)约为 10^{-24} 厘米。在暴胀 10^{50} 倍后，同一个区域就会大到 10^{26} 厘米，全是由那个几乎完全均匀的小时空点变来的。[8] 这样一个暴胀宇宙的尺度是

1 亿光年。但是在暴胀结束即 $t=0$ 之后将近 10^{-30} 秒时，这个暴胀宇宙里那个后来增大成我们今日可见宇宙的空间区域还只有 10 厘米。我们所能看见或知道的一切，都只是在暴胀宇宙里的一个泡中。

暴胀方案的许多详情都有待研究。如里斯(Martin Rees)在日内瓦的 ESO/CERN 会议上指出的，今天的游戏状况，颇像伽莫夫在 20 世纪 40 年代搞标准大爆炸模型时的状况。指望所有的答案都一下子掉出来是不现实的。宇宙学家正在两个主要方向上努力。一个是试图解释极早期宇宙中的量子涨落怎么能在创世后几分钟时增长成为星系的种子。从引力的角度讲，还有关于星系在宇宙中究竟有多重要的有趣问题。这是因为，按照对可见星系计数基础上的密度估计，使宇宙平坦所需的物质只有不超过 20%是以星系的形式存在。标准大爆炸模型要求宇宙中只有三种(或最多四种)轻子族，由模型也能计算出宇宙中应有的重子数，结果与对星系的计数符合得相当好。暴胀模型说时空必定平坦，那么所需物质的其他 80%(或更多)是在哪里？可能是规范粒子形式的非重子物质，或是各种微子，或是粒子理论的其他什么古怪副产品。好戏还在后头呢。天文学家用计算机模拟了宇宙中星系的形成过程，也就是看引力如何随着宇宙的老化和膨胀而把物质团块拉到一起，他们发现，要与观测到的物质成团性相符，在超星系团之间的空洞里就得有大量寒冷而黑暗的物质。这种冷暗物质里的粒子每个都有 1GeV 或更大的质量，它们组成了宇宙中引力质量的主体。它们不是重子。但正是这些粒子的引力解释了为什么组成可见恒星和星系(还有我们自己)的重子是以我们所见到的方式分布。

暴胀对认识今日宇宙及其终极命运有着迷人的潜力，因而在今后的年头里会继续使宇宙学家忙碌，使书架被塞满。但是那个研究方向与我在本书里所遵循的正相反，我虽不情愿也只得再转回到寻求对宇

宙起源认识的主线，讲述1984年才出现的有关暴胀的另一项新进展，它把时间进一步推前，从10^{-35}秒时的GUT破裂希格斯相变推前到量子引力刚刚变得不再重要的时候，也就是创世后仅仅10^{-43}秒之时。

原初混沌与终极秩序

尽管古思的工作代表了观念上的突破，在他的模型和新暴胀方案里居于中心地位的希格斯场却仍是一个薄弱环节。新暴胀方案之所以有效，是因为对希格斯场的参数作了很精细的调节，而之所以允许“微调”又是因为我们并不真正知道希格斯场的行为，从而能为自己辩解。这套技巧也只有事先把场参数恰当给定才管用，因为我们知道想要得到的“答案”，[9] 而理论的创立者们自己也认为这样做不合理。他们所希望的是实在的大统一理论，在其中平稳的转变能自然地发生并自动地给出“正确答案”，而用不着任何微调。但由于有关细节与粒子物理密切相关，而对粒子物理的认识又不够好，所以直到1985年这仍只是一个良好的希望。有些理论家却无所畏惧，他们解决困难的尝试是，从一个本就知之甚少的理论（GUT暴胀）跳跃到另一个知之更少的理论（SUSY暴胀）。

欧洲核子研究中心那些已变得对宇宙学大感兴趣的粒子物理学家就是这么干的，他们避开对GUT时代结束时对称破缺希格斯场微调的困难，回溯到10^{-43}秒时使引力与GUT力分离的对称性破缺。这也就是在GUT破缺之前先暴胀，而宇宙成为由辐射、夸克和神秘的冷暗物质粒子所充满的空间则被留作后来的事情。欧洲核子研究中心这个项目组的主要发言人之一是希腊粒子物理学家纳诺普罗斯（Dimitri Nanopoulos），他把这个新主意叫做“原初暴胀”，并把据此建立的各

种不同模型统称为 PRIMO。他和同事们本希望用 N=8 超引力理论来描述极早期宇宙的暴胀过程，但目前对他们来说这太复杂了。于是他们就用了比较简单的 N=1 超引力理论，希望要么以后再把所得结果推广到 N=8，要么能找到把 N=8 理论简化成 N=1 的途径。这可是一项很大的工作，在这里对它作细述未免就显得太傻了，因为在本书印出来之前它可能已被新进展推翻了。他们迄今的绝大部分工作的确也只是以 CERN 预印本的形式发表的——“只是”，出于担心冒犯了我在欧洲核子研究中心的那些朋友，我要指出茨威格也是以同样形式把他的夸克(爱司)模型介绍给全世界的。但这项研究有一个重要特点，它所对付的是物理学所能描述的终极能量和密度。引力是准备包括进统一理论的最后一种力，又是在大爆炸中分化出来的第一种力。确确实实再也乞求不到什么别的东西来提供驱动暴胀的机制，也再也没有更早的时间来往里面塞进点什么。所以此类研究必将终结对宇宙本质的揭示，哪怕所揭示出来的东西并非如我们所预期的也罢。

另一种探索也推回到创世时刻，而不必顾及希格斯场(这些场有着很特别的假设性质，从来没有被直接观测到过)，林德继其他苏联研究者之后正在进行这种探索。这条路线为许多苏联宇宙学家所赞同，但在苏联以外就不如新暴胀那样广为人知。林德抛弃了高温相变推动宇宙指数式暴胀的思想，而来探究宇宙诞生时就充满其间的所有标量场怎样落到稳定的最低能态上。他认为，在创世时刻每种场都以混沌状态产生出来，在宇宙的不同部分有着不同的值。场会试图“滚落”(用碗中弹子的比拟)到最低能态，但滚落的方式却取决于它产生时的状态。在宇宙中场几乎均匀又远离平衡态即最低能态的那些区域，开始时场滚落得很慢，就像新暴胀模型里缓慢滚出高原的场那样。但尽管场丢失能量很慢，该膨胀宇宙中辐射和粒子的能量密度却

随着宇宙膨胀而迅速降低，并很快降到与缓慢滚落的场的能量密度相当。这时场能量支配了膨胀，而按照爱因斯坦方程，只要能量密度恒定或者只是很缓慢地变化，宇宙就必定作指数式膨胀。换句话说，单是任何一个标量场朝其最低态的缓慢滚落就足以激发暴胀，并不要求相变或“超冷却”。这些思想中有些是与欧洲核子研究中心的PRIMO模型共同的，并且肯定比由古思首先提出的希格斯机制更简单也更自然。

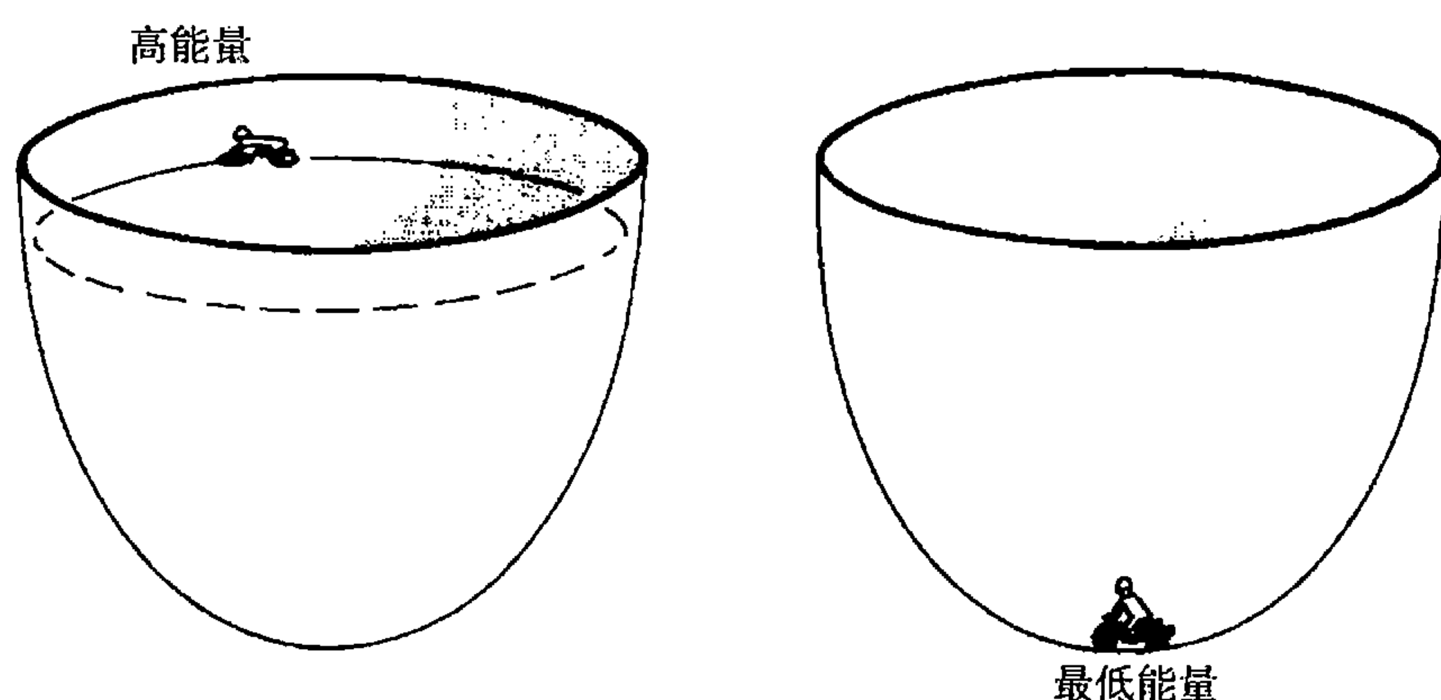


图 9.7　宇宙的能量可以用走壁的摩托车来比拟。车子开得快也就是能量大时可以边旋转边升得很高，这相当于宇宙创生时的高能量密度状态。车子减速也就是能量变小时就不可避免地落向最低点。但宇宙在冷却和能量密度减小时却有不止一个最低点可选择（见图 9.9）。

一旦场达到了其最低值，那一直在驱动暴胀的条件就不再存在了。场会在最低点附近来回振荡，就像弹子滚落到碗底时也会来回振荡一样，在这个过程中场的能量都转化成粒子对。随着动能被摩擦转化成热，弹子会慢慢地落定在碗底的最低能态；振荡的场也会慢慢落定在最低能态，其“滚动”能量就转化为成对产生的粒子。与以前的暴胀方案一样，宇宙被再加热到接近于 10^{27}K 并充满了 X 玻色子等粒子。然后它就以标准模型的更稳重的方式膨胀，并迅速冷却。在创世时刻离平衡态最远的区域经指数增长就成为最大的宇宙“泡”，它可能已经暴胀了 $10^{1000000}$倍，这可是个使人脑子转不过来的数字。如果

这幅图像正确，那么我们今天在自己周围所看到的秩序就是由原初混沌经暴胀而产生出来的。

林德把他这个最新版本的暴胀称为“混沌暴胀”，并提出这个概念可能还会与卡鲁查—克莱因宇宙学有关联。如果时空真是从 11 维开始，那么在不同的宇宙泡即不同的“领域”里，某些时空维度的紧缩和滚动过程就可能以不同方式发生。我们这个领域正碰上是 7 个维度被隐藏了，还有 4 个显露着。其他理论家还建议，正是那 7 个原初维度的紧缩和滚动推动了宇宙的最初膨胀。但所有这些想法，包括林德的在内，都至少也像欧洲核子研究中心小组的 PRIMO 那样是思辨性的。

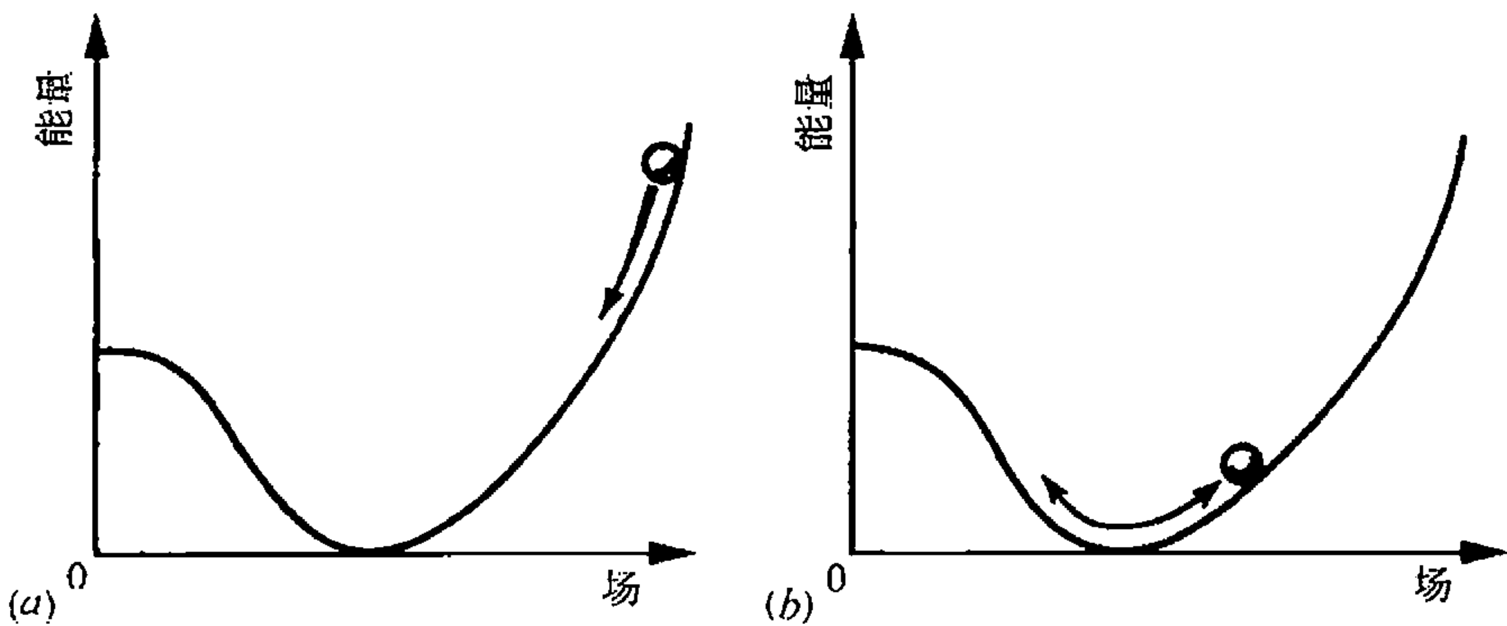


图 9.8　在宇宙“落入”最低能态的过程中，它的场能量会来回振荡一段时间，就像在碗里下落的弹子的振荡一样。正是在这种振荡中剩余的场能量转化成为成对产生的粒子。

以暴胀为主题的变奏曲之纷纷出现，既表明大门打开之后呈现出了丰富多样的可能性，也反映了宇宙学家和粒子物理学家对这个基本思想的重视。在他们看来已经建立得足够好并已显示出足够强的效力、因而能像 40 年前伽莫夫接受大爆炸思想那样信心十足地接受的这一基本思想，就是很早以前宇宙确曾有过一个暴胀阶段，正是暴胀使一粒不大于普朗克长度即 10^{-33} 厘米的种子长成我们所能看到的一

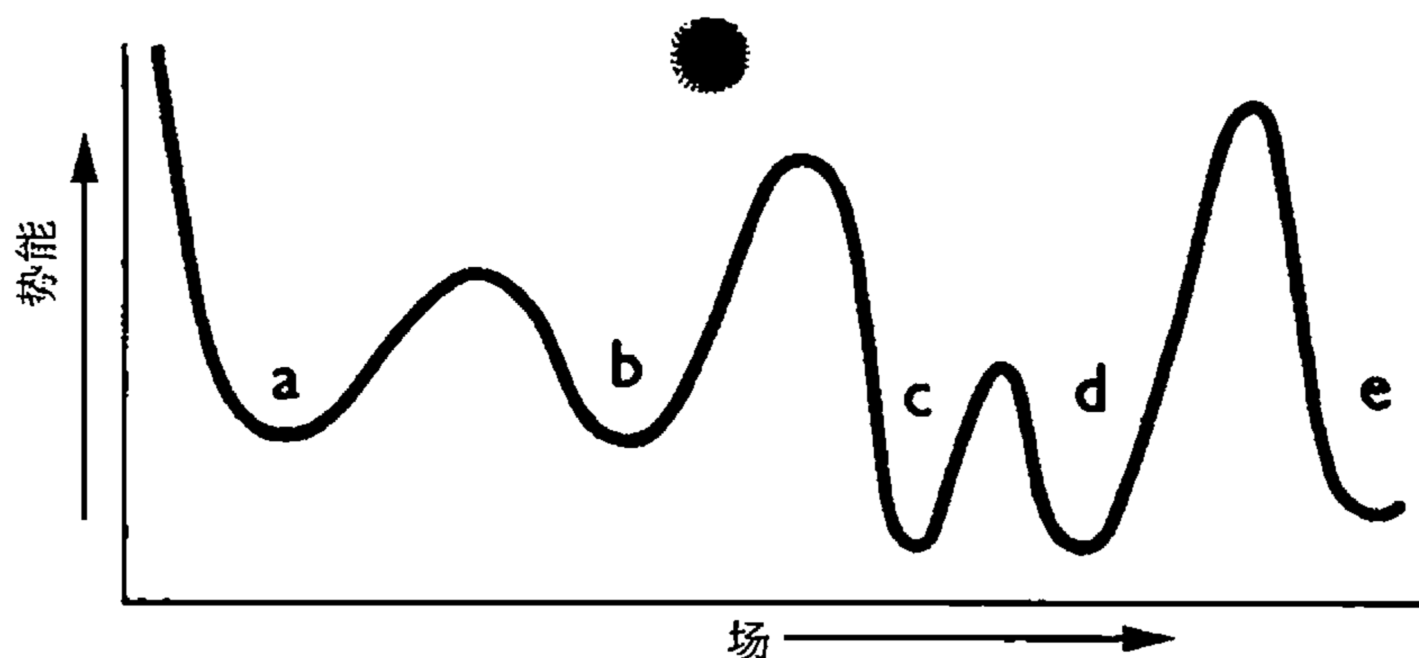


图 9.9　按照林德的"混沌暴胀"方案，宇宙可以选择所"落进"的最低能态。这个图虽过于简化，却仍能表明最初提出的极简单的暴胀方案怎样能改得更为复杂，以与真实世界符合得更好。

切。但是要说层出不穷的各种详细方案中的哪一个（如果有那么一个的话）会被最终证明反映了真实世界，实在还为时太早。我想还是让这个领域的两位开拓者古思和林德自己来说为好。林德很相信正在朝正确方向取得进展。他告诉我："老方案死了，新方案老了，而混沌方案状况良好。"古思曾经认为宇宙学是一个你说什么都行、没人能证明你错的领域，现在的观点则稍有不同。他说："现在看来很容易证明一个宇宙学方案是错的，而要建立一幅完全自洽的图像就比我原来设想的困难得多。"

也许暴胀模型还没有做到"完全自洽"。但它们已经绘出一幅十分重要的有效的图像，一幅创世时刻的图像，我们所知的整个宇宙那时还被装在普朗克长度的范围内。这就使得对创世时刻本身作出数学描述终于有了可能。

注　释：

1. 致罗伯特·胡克（Robert Hooke）的信，见 *Oxford Dictionary of Quotations*，p.362。

2. 为什么与李政道和杨振宁相比获奖有所拖延呢？因为到了 20 世纪 60 年代中期，发

现某一守恒“规则”被违反已不再像在1956年那样是一种巨大的冲击了。

3. 萨哈罗夫1921年生于莫斯科，1942年毕业于国立莫斯科大学，1945年进入莫斯科的列别捷夫研究所，像他父亲一样成了一名物理学家。他对苏联热核武器的发展起了重要作用，1953年32岁时成为苏联科学院有史以来最年轻的院士。他还是社会主义劳动英雄，获得了列宁勋章和许多其他荣誉。但到了20世纪50年代后期他变得热心于社会活动了。

开始时他的公开评论是关于苏联教育体制的改革，发表在《真理报》上，其中一些在20世纪60年代初还成了官方政策。60年代时萨哈罗夫的科学研究集中于宇宙学和大爆炸的初始阶段。1968年他发表了一篇呼吁裁减核武器的文章。他在70年代越来越多的政治活动使他被放逐到高尔基市。当时他的科学事业似乎完结了。但是后来宇宙学的新进展回应了萨哈罗夫在60年代的一些理论工作，于是在1984年他讨论与宇宙的创生和早期演化有关的基本概念的文章又出现在苏联的《实验与理论物理学杂志》上。

4. GUT是20世纪70年代才提出的，萨哈罗夫在此之前就能以其深刻的洞察力来提出使今日宇宙中有物质存在所必须满足的要求，这肯定值得授予诺贝尔奖，当然也可能很快会授予。或许也会很快给茨威格授奖，因为诺贝尔奖委员会还没有对夸克的存在作出过表示。让他俩分享倒是个好主意！

5. 引自古思的文章，收录于G. W. Gibbons, S. W. Hawking & S. T. C. Siklos(eds.), *The Very Early Universe*, p.201。

6. *Physical Review*, Vol. D23(Jan., 1981), p.347.

7. 10^2是100即10×10，2^2是4即2×2，2^{100}就是2与2相乘100次。另一种写法是$(2^3)^{33}$，2^3是8，所以也就是8^{33}，即8乘8共33次。如果只是要对指数暴胀的意思有个粗略了解，就可以说8与10足够接近，因而2^{100}足够接近于10^{33}。同样道理，暴胀10^{50}倍就大致相当于加倍150次，也就是增大到2^{150}倍，以每10^{-34}秒加倍一次的速率，这只要用150×10^{-34}秒，即1.5×10^{-32}秒就可完成。

8. 1光年是9.5×10^{17}厘米，但宇宙的膨胀能使时空延伸得远“快于光”，因为并没有任何东西在穿越时空运动。所以由古思和其他宇宙学家提出的宇宙指数型暴胀，确实可以使一个比质子小很多很多的时空区域在远不到1秒的时间里增大到约1亿光年的尺度。

9. 重正化的影子！

第十章　创世时刻

我们所知道的宇宙是怎么来的，这是自有宗教和形而上学以来这两个领域里的一个主要话题。它也是物理学家的难题，即使在大爆炸的思想出现之前就已为众多学者所探讨。对这个问题的最好表述，也就是最深刻最基本方式的表述，还是最初的那个奥伯斯佯谬：为什么一个寒冷而黑暗的宇宙里却有着高温的明亮恒星?

所有物理系统的运作都是使热从高温物体流向低温物体。在这个过程中系统会变得更无秩序，最后达到的稳定状态具有均一的温度，其中的所有粒子都做无规运动。无论温度是多高，高到如大爆炸的火球，或者是多低，低到如今天的背景辐射，都不重要。重要的是温度必须均一，系统必须不包含任何结构，也就是没有任何种类的“信息”。

这样一个温度均一的平滑而无特征的系统，就被说成是具有最大可能的熵，熵或者说无序性必定总是增大，这是由实验所揭示由理论所导出的一条自然界基本定律。今天地球表面上的生物能够获取太阳

光的能量，在局部区域造出结构和秩序而使熵减小，但在这样的情况下该定律只是被扭曲，而不是被打破，因为这样做总是以别处的熵增大为代价，而增大量比局部减小量更大。太阳的核燃烧就对应着熵的巨大增长，相比之下地球上生物造成的熵减小实在是微不足道。地球上的工程师建造了一座桥、一架飞机或一栋楼房，他们是局部地增加了秩序和减小了熵，但代价却是在出产原材料的地方和把原材料运到制造厂时增加了无序和增大了熵。只有在热以某种途径从高温物体流到低温物体时才能做功。熵量度可用于做功的热，熵越大就表示可利用的热越少，也就是说由温度差驱动的由高温物体到低温物体的热流越小，这种热流是所有功和动力的终极源泉，熵越大就意味着做功的能力越小。如果温度到处都一样，那么即使它很高，也没有热流，也就不能做功。但做功又从来不能百分之百地有效。东西总是要被用坏的，这是对熵的不可抗拒增长的最简单表述，正式的名称则是热力学第二定律。但宇宙是在低熵和不平衡的状态，这是由恒星向空中倾泻能量的事实显示的。宇宙还远没有“用坏”，它就像一架大钟，很久以前被上紧了发条，今天还在走着。那么发条是怎么上紧的呢？宇宙怎么能在开始时有这么小的熵、这么好的秩序，以致大爆炸之后 150 亿年我们还能看到黑暗天空里的明亮星星呢？

时间的起源

我们现在对热力学定律和熵在物理系统行为的作用之认识，得自于玻尔兹曼(Ludwig Boltzmann)在 19 世纪的工作。玻尔兹曼是奥地利人，1844 年生于维也纳。他对科学的诸多贡献之一是得出了对气体达到平衡倾向的数学描述，处于平衡状态的气体有均一的温度，气体分子做无规运动。正是玻尔兹曼阐明了熵是系统无序性的量度，而

熵这个概念是由德国物理学家克劳修斯(Rudolf Clausius)于1865年提出的。

玻尔兹曼用的是统计方法。他发现，气体的行为可以用巨大数量分子的无规运动来解释，这些分子互相碰撞，也与盛装气体的容器的壁碰撞。描述这种无规相互作用的大量粒子系统的规则非常精确可靠。这些规则从简单的概率计算开始(就像计算把一对骰子连掷三次得到两个六点的概率)，能给出气体分子系统取各种状态的可能性大小，并指明该系统可能性最大的状态是什么。由玻尔兹曼提出并由后人改进的一组方程式就称为统计力学。这门学科极为成功，堪称是现代物理学大厦的一根顶梁柱。但它也给出了一些有趣的和超越常识的预言。

我在其中工作的这个房间实际上是一个充满气体的盒子，房间里的所有空气分子都在无规地运动。这就是说有可能(尽管可能性很小)所有这些分子都突然运动到墙角里，使我无从呼吸。或者举另外一个例子。我放一块冰到饮料里，冰会受热而溶化，饮料则因把热给了冰而冷却。但是按照玻尔兹曼方程，这只是一个统计规律。也有这样一种虽然很小但毕竟存在的可能性，有一天我把冰放进饮料后惊讶地看到，冰变得更冷，而其周围的液体开始沸腾。在物理定律允许的范围内熵减小是可以想象的，但这样的事发生的可能性却小到几乎成了不可能。

要了解究竟是怎么不可能，不妨考虑一下得有多少分子一起行动才能使杯中的饮料自动沸腾，当然饮料得从周围的空气中或从就便的冰块中吸取热量，空气和冰就会冷却。1克氢气里的原子数是6×10^{23}，即6后面跟了23个零。氧原子重量是氢原子的16倍，每个水分子含有1个氧原子和2个氢原子。作为粗略计算可以忽略氢原子的

重量。16 克水也含有约 6×10^{23} 个分子，假定我所喝的只是水，那么 10^{24} 个分子只是很少一点水，大约是 1/30 品脱(大约 16 毫升)。按照法国数学家庞加莱(Henri Poincaré，生于 1854 年，卒于 1912 年)所得出的严格统计规则，所有这些分子“一起行动”来使熵减小的概率依赖于分子数。这个概率是 $1/10^n$，其中 n 是所包含的分子数。其实杯子周围的空气分子也得参与这个与正常的熵流动相反的过程，但现在姑且放在一边。即便如此，杯中的水自动沸腾的概率也只有 $1/10^{10^{24}}$，分母是 1 后面跟着 10^{24} 个零。作为比较，宇宙年龄以秒为单位约是 5×10^{17}，即 5 后面只有 17 个零。千万别去打赌说冰块放到饮料里除了溶化还能干别的了！

这种古怪的违反常识的事情发生的概率实在太小，非但不会影响我们的日常生活，甚至也不会影响物理学家所能想得出来的任何奇特实验。但这种事情与整个宇宙会有什么关联吗？玻尔兹曼认为可能有，那是在他为我们人类怎么能出现在这么一个有秩序的低熵宇宙里而苦思冥想的时候。他推测宇宙的“真实”状态必定是一种热平衡态，是一片温度恒定的无序的均匀海洋，已经永久存在还将继续永久存在，因为在没有温度差从而没有热流也不能做功的情况下时间根本就没有意义。[1] 如果什么都不改变，时间也就不存在，一切永存不朽。但在这个无限的永恒宇宙里会有无规的涨落过程，在很有限的时间里造出缺少熵的区域，正如一个偶然涨落会使我杯子里的水沸腾那样。或许我们的整个宇宙就是这么一个偏离平衡态的统计涨落。这样一种涨落所发生的机会也与可见宇宙中的原子数相关，大约是每 $10^{10^{80}}$ 年里有一次。

这简直是完全不可能。但要知道，我们所谈的是无限长久地存在的总宇宙，其中的这么一次涨落毕竟只是局部反常事件，总有用不完

的时间来等待它发生。或者也可以想象总宇宙是无限广阔的，于是任何可能的事情，哪怕可能性再小，也一定会在其中的什么地方发生。[2] 不管那个均匀的总宇宙是怎么等了 1 后面跟 10^{80} 个零这么多年，总之是有了这么一次机会，在总宇宙中像我们宇宙这么大的一块地方里，所有的热都自动流进了恒星，使它们的温度升高到几百万度。那么我们今天所看到的一切，就简单地只是朝着平衡态的复归，平衡才是我们宇宙的真实而长久的状态。

玻尔兹曼的思想以其最初的形式不可能走得很远，部分是由于难以解释为什么宇宙必须如此均匀，而为什么在这么大的时空范围里熵又已经以同样方式减小了。我们又回到了视界和平坦问题。但是由统计起伏来创造宇宙的思想在 20 世纪 70 年代以一种不同的形式复活了，并且在 80 年代当暴胀方案受到重视时也突然变得像是很有道理。

量子宇宙

随着返回创世时刻的探索，老思想不断以新装束出现，这倒是颇为使人惊奇的。玻尔兹曼的宇宙图像，或者说他的模型，真正就是一个“稳恒态”模型，其整体特征(或者说缺乏特征)是永恒和不变，我们看到周围进行的活动都只是短暂的起伏。最简单的现代形式的稳恒态模型是在 1948 年提出的，已经早就被抛弃了，因为我们能看到宇宙随着从超密态膨胀开来的演化和改变。但在 20 世纪 60 年代，作为与进展中的大爆炸模型的殊死一战，霍伊尔和他的印度同行纳里卡(Jayant Narlikar)提出了一个新版本。这回是一个高温度高密度的永久的稳恒态宇宙，其中有时出现一些在胀大的泡，就成为像我们宇宙这样的膨胀着演化着的时空区域。这的确有玻尔兹曼的影子，但也未

尝不可看作现在这么时兴的暴胀模型的一次预演。

为了提供暴胀的驱动力，霍伊尔和纳里卡不得不发明出一种新的场，并取名为C场，因为他们预定要把它用于物质的创生。*于是他们在20多年前就有了这样一个宇宙模型，即一种很密集的状态在一种场的能量驱动下暴发式地迅速膨胀，那种场还生成粒子。从概念上讲这与现在时兴的宇宙暴胀理论没有什么差别，后者说的也是一种或多种场(也许是希格斯场，也许不是)的作用使宇宙从高密态暴胀出来，也是场以粒子的形式释放能量。1984年，霍伊尔和纳里卡各自撰文，强调他们的老C场模型与暴胀之间的相似。他们希望其他宇宙学家注意的要点是，像C场这样的暴胀过程能在稳恒态背景上相当容易地启动，正如玻尔兹曼涨落能从混沌中造出秩序一样。坦率地讲，物理学家们并不乐意C场思想复兴，任何稳恒态理论的变种都被认为不合潮流。但是该肯定的就得肯定，C场确实使人联想起暴胀(霍伊尔和纳里卡会说是暴胀使人联想起C场)，而重要的是记住这样一种可能性，在我们所称的创世时刻并不真有什么奇点，并不真有什么无限高密度和温度的状态，有的只是一个我们称为宇宙的局部时空区域从一种状态到另一种状态的转变。

但1985年形式的暴胀理论作出的最惊人的引申却是，整个宇宙可能确实是从虚无中来的，是作为一种量子涨落而创生的，正如量子不确定性允许虚粒子对出现并短时间存在然后湮灭一样。

这个思想首先于1973年12月出现在《自然》上的一篇论文里，作者是纽约城市大学亨特学院的特赖恩(Edward Tryon)。特赖恩提出了他称为可以想象得到的“最简单最有吸引力”的大爆炸模型，即

* C是英文创造(creation)一词的首字母。——译者

"我们的宇宙是一次真空涨落"。[3] 他以这样一个模型来参与宇宙学争论的出发点是，计算表明任何闭合宇宙的净能量必定为零。

粗略地讲，这可以由宇宙所具有的负引力能来理解，这个引力能很大(负值)，能够抵消宇宙中所有物质的质量能量 mc^2。这是引力的一种很奇怪的特征带来的结果，以至于物理学家几乎没有公开承认过。如果用引力方程来描述一个物质集合的引力能，就会发现唯一对应于零能量因而可以由此来作测量的有意义的状态是，所有物质都散开到无穷远处。无论是拿原子、或行星、或恒星、或星系作为基本物质单元来考虑都是这样。当物质单元分离得尽可能想象的那么远时能量为零。现在怪事来了。物质在引力作用下落到一起时会放出能量，于是引力势能减小，由于它是从零开始的，因而就会变成负值。所以真实宇宙中的任何物体，比如一颗行星，既然没有被散开到无穷远处，就必须从负能量开始，而且如果它收缩，它就释放能量，它自己的引力势能就变得更负。这可不只是一无所获，而是比一无所获还要差。按照牛顿引力理论，行星的势能可以有多负是没有限度的。每次它一收缩，能量就释放，引力能就变得更负。

用数学术语讲，在牛顿物理学里能量状态没有"下限"，这是把分散物质的能量定为零的一条理由。广义相对论则对质量为 m 的物体所具有的负引力能量给定了一个限度。如果质量 m 能够全部集中空间一个点上，即成为一个奇点，那么相应的负引力势能是 $-mc^2$，与爱因斯坦质能数值相等符号相反。但这只是一个更精细更严密的数学论证的粗略表述，对那个数学论证这里不能讲述，但它的确证明，一个闭合宇宙的总体能量为零。[4] 这是把古思的"免费午餐"推到了合理的极端。如果宇宙包含的能量本来是零，免费也就不足为怪。不是什么一无所获，而是本来就没有付出。特赖恩指出，不确定性关系

$\Delta E \Delta t = h$ 允许任何能量为零的东西存在任意长的时间，因为如果能量为零，则能量的不确定值 ΔE 也为零。从真空中“借取”能量来创造宇宙不会有问题，因为在开始时并不需要有任何整体能量，也不必忙着把借的能量归还，反正算起账来不会有任何损失！

对不确定法则的这种朴素的过于简化的解释在 20 世纪 70 年代没有引起多大反响。它只是那种物理学家们在喝咖啡时随便谈谈就过去了的话题，它显然也没能对我们的宇宙作出精确的描述。例如，如果严格地由虚粒子对的产生来类推，就要求我们的宇宙包括精确等量的物质和反物质，实际情况看来并非如此。而且整个论证的前提是宇宙闭合，但在 70 年代中期宇宙学家的广泛共识却是宇宙开放。最后，如果包含了我们宇宙全部质量的量子涨落是产生出一个超密态，为什么它不在自身引力作用下立即收缩成为一个奇点呢?

这些困难看来是无法克服的。但是暴胀的出现改变了一切。暴胀要求必须有足够的暗物质来使宇宙非常接近于平坦，接近到我们根本分辨不出其曲率，而暴胀并不“在意”宇宙究竟是落在平坦分界线的闭合一侧还是开放一侧。按照 X 玻色子衰变的 GUT 描述，含有等量 X 和反 X 粒子的高能原初宇宙会在适当时候(“适当时候”是大约 10^{-35} 秒时)演化成正物质稍有剩余的宇宙。于是特赖恩在 20 世纪 80 年代借助于暴胀来重提自己的思想就不足为怪了，塔夫茨大学的维连金(Alexander Vilenkin)在 1982 年也是这样做的。[5] 但维连金比起 1973 年的特赖恩来前进了一步。特赖恩讲的“真空涨落”的意思是在宇宙出现之前就有某种形式的时空度规存在；而维连金却试图建立一个模型，其中空间、时间和物质全都是作为虚无中的一次量子涨落而从确实的虚无中创生出来的。他在自己的一篇论文(《物理学快报》，1982 年 11 月 4 日第 117B 卷第 26 页)中写道：“宇宙从虚无中创生的

概念是一种狂想"，但他接着论证，这个概念如何在数学上是等同于正负电子对的产生和随后的湮灭，而这个过程又如何等同于一个电子从虚无中产生，顺着时间前进一会儿，然后转过来逆着时间退回，直到遇上它自己的产生。在这篇和其他一些最近的论文里，维连金对特赖恩在20世纪70年代初的猜测性的想法作了许多如此这般的高雅数学加工。

对特赖恩来说，他的猜想复活自然事关重大，这尤其是因为他在宇宙学研究中是一位孤独者。他于1940年出生在印第安纳州的特雷霍特，1962年在康奈尔大学得到物理学学士学位，然后到了加利福尼亚大学伯克利分校，在那里为温伯格的量子场论和广义相对论课程所吸引，并很有运气(当然也很有才能)，成了温伯格的博士生之一。完成学位论文后他到了哥伦比亚大学，搞 π 介子对之间碰撞散射幅度的计算，得了一个"π 介子特赖恩"的绰号。他从来不跟任何人合作，所有论文都是以自己为单一作者发表，这恐怕在他那一代粒子理论家里是独一无二的。从儿时就有的对宇宙学的爱好使他开始推测净能量为零的闭合宇宙的可能性，但他曾很有兴味地回忆过20世纪60年代末曾使他窘迫的一件小事。

宇宙学家席阿玛(Dennis Sciama)从英国来访问哥伦比亚大学，作了一个关于宇宙理论最新进展的报告。就在席阿玛稍作停顿的时候，特赖恩脱口而出："也许宇宙是一次真空涨落！"这使他自己和所有其他人一样都吃了一惊。随之而来的笑声过后，这位以粒子物理而非宇宙学为专业的年轻研究者自己也把这事给忘了。直到他在《自然》上的文章发表之后才由一位同事对他提起，他才明白自己原来是被潜意识推动去完成这项工作，而自己的记忆已在脑子里埋藏了三年。

特赖恩于1971年来到亨特学院，1972年中期那个宇宙是量子涨

落的念头在他脑中又突然闪现，这回他就完整地写下来了。潜意识一直在推动他阅读宇宙学的文献，并寻求对那次使他发窘的大笑的回答，直到回答得到了才显现在他脑子里。他的文章在《自然》上发表后很快就有约150位同行要求得到复印件，但这个思想随之又被冷落，直到暴胀概念使它被重新记起。一个曾被嘲笑的老思想又变得时兴并被看作概念上的重大进展，对此该作何感想呢？在我这样问时特赖恩说："所有优秀的科学家都是幻想家，他们幻想着发现有重大意义的未知现象……很难夸大他们从中得到的满足。"

这个思想当然仍是高度推测性的，但现在已经有吸引力得多了。它的最大优点是，不再需要在创世时刻就产生出今日宇宙中的全部物质。现在所需要的只是产生出一个闭合的时空区域和能量，一个比质子还小得多的自含的微宇宙，它的温度并不高，也只是稍有膨胀的倾向。如果没有暴胀，这样一个微宇宙会很快坍缩。但是特赖恩说，有了暴胀（或者如他所称的"冷大猛冲"），就能把一个时空小点胀到巨大的尺度，暴胀结束时标量场能量转变成X粒子对和其他粒子，从而出现一次创造高潮，并在 $t=10^{-35}$ 秒时造成热大爆炸本身。引力一直在使膨胀减慢，宇宙最终会停止膨胀并转过头来，在遥远的将来又坍缩回到一个奇点。时空本身以及其中包容的一切都将消失在那个奇点里。再也没有任何迹象显示我们的宇宙曾经存在过。这个方案虽仍是推测性的，但至少已能比较简洁地符合暴胀宇宙学的整体框架。如果真有一个创世时刻，那么作为量子物理学最奇特最基本的特征之一的量子不确定性概念，看来最有希望对宇宙怎么来的作出解释。倘若果真如此，就可能确有一个标志着宇宙的时间起始边界的创世时刻。但还有另一种观点，它甚至更深地植根于量子物理学的基础。剑桥大学的霍金定义了一个描述宇宙整体的量子力学波函数，并像对量子物理

学的任何其他波函数能做的那样用路径积分来处理。他认为，宇宙即使在创生时刻可能也没有边界。

奇点探索者

霍金是那种罕见的学者之一，他们的工作能牢牢吸引公众的想象力，并且常是报纸和杂志报道的题材。这部分地是由于他研究的课题扣动了我们所有人的心弦——黑洞、时空奇点，还有宇宙起源的奥秘。但还由于他得了一种使人严重残疾的病，叫做运动神经元病。演员尼文(David Niven)就死于这种病；它损害人体的神经和肌肉功能，严重时使患者不能行走，讲话也极为困难，霍金现在就是这样。一位残疾天才与不可抗拒的命运搏斗，得到远高于我们绝大多数人的对宇宙的认识，还得克服障碍把这些认识传递给他的同行，这显然是报纸报道的好材料。但有时对霍金生活的这一面宣传太多了。他的身体肯定是有了残疾，但他的大脑和智力完全没有受疾病影响，他的科学成就是第一流的。值得惊讶的倒不是一位残疾人能在对宇宙的认识上取得如此进展，而是竟能有人取得如此进展。霍金自己曾说他很幸运，能有一个只靠大脑来思考就行的职业，而身体障碍虽然严重，在他看来却只是一个次要的痛苦。身体问题对他工作以外的生活当然会有很大影响，但这与他的科学发现的关系，大约同爱因斯坦拉小提琴与探索大爆炸的故事的关系差不多。

霍金生于 1942 年 1 月 8 日，他喜欢对人说，那一天正好是伽利略去世整整三百年。他父亲在国立医学研究所研究热带病，鼓励他也走学术道路，第一个目标是要上牛津大学。但在霍金决定要学数学时他父亲可就不鼓励了，而是劝他改变主意，理由是数学家找不到工作。但霍金在 1959 年还是进了牛津学数学和物理。他的同学和老师

都记得他是个优秀的学生，脑子与别人都不同。他轻而易举地通过了所有考试，以最优成绩完成了学士课程，接着到剑桥大学开始研究宇宙学。

那是20世纪60年代初，霍金开始同奇点打交道，他的所有重大科学贡献都围绕着这个问题，这个问题也是认识创世时刻的关键。他过去和现在都迷恋于数学奇点的概念，不仅物质而且时间和空间都要么被挤压进这个点而不再存在，要么像在大爆炸的情况那样从这个点里产生出来。标准的相对论方程预言了奇点的存在，但在20世纪60年代初几乎没有谁认真看待这个预言。奇点被认为只是这样一种暗示，即以物质在时空中均匀分布为前提的最简单形式的爱因斯坦理论不适于描述混乱的超密状态，而对方程式的更好认识可能会是，当坍缩物体向奇点趋近到某种程度时会出现“反弹”而使它重新膨胀，或者有什么别的效应使坍缩停止而不致形成无限大密度的点。要么这种想法是对的，要么就是爱因斯坦理论本身不完备，在极高密度也就是极强引力场的情况会失效。霍金决定搞清楚这个问题。但还要过好几年这个决定才会产生结果，因为就在他当研究生的第一年即1962年，他的病首次出现了症状并得到了诊断。他被告知只能再活几年了，于是他变得沮丧、酗酒，研究也停止了。

但几个月后霍金的病情停止发展并稳定下来。他的身体已轻度残疾，但没有变得更糟。同时他首次认识到自己的智力完全没有受疾病影响，而且不管身体怎样都不会受影响，自己的工作又完全是脑力的，不管身体条件怎么恶化都能进行。据像我这样熟不拘礼的朋友所知，从此以后霍金无论在个人生活上还是在工作上都再也没有停止不前。他于1967年结婚，有两个儿子和一个女儿，尽其所能地过正常生活。也是在20世纪60年代后期他的研究工作开始得到了重视。[6]

这个时期霍金的主要成果之一是与当时在伦敦大学工作的数学家彭罗斯(Roger Penrose)合作取得的。他们证明，经典形式(即不考虑量子效应)的广义相对论方程绝对要求在宇宙诞生时有一个奇点，即时间开始的点。在经典广义相对论的框架里无法绕开奇点问题。要在真实宇宙中避免奇点，唯一的希望是改进相对论，即引入量子效应并建立量子引力理论。20 世纪 70 年代，霍金在对黑洞的研究中引入量子效应，得出了惊人的结论，即黑洞能够“蒸发”并且最后还会爆炸。这项工作使他为公众所注目，至少是在科学杂志上。1974 年他还只有 32 岁，就已被选为皇家学会会员。

那时他已经由于病情进一步恶化而被限制在轮椅上了。但在过去的 10 年里，他的身体状况却似乎总保持着那个样子。他对自己身上的肌肉只有非常有限的控制，是倒在而不是坐在轮椅里，讲话很吃力，不是很了解他很熟悉他声音的人几乎什么都听不懂。他所得到的荣誉包括 1978 年的爱因斯坦奖[7]，1980 年他成为剑桥大学的卢卡斯数学教授，以前享有这个职位的人中有狄拉克和牛顿。这些荣誉，还有世界各国的大学授给他的一大堆名誉学位，通常都属于这样一些学者，他们已经完成了最重要的研究工作，现在坐在舒适的显赫的位置上，充当行政官员和导师。实际上很少有数学家能在 30 岁后在新的研究领域取得大成果；我们也常听说，新思想是来自不受常规限制的年轻人的头脑。但是霍金的头脑总是那么锐利，他现在提出了一个宇宙模型，试图把广义相对论和量子物理结合在一起，不仅将去掉使人不舒服的创世时刻的奇点，而且原则上能一股脑儿解释一切。也许对他的工作来说残疾真有好处，因为他不能去当什么委员会成员或什么挂名首脑，而只能继续他唯一能做也已经做了 20 年的工作，就是在脑子里构造对世界的新数学描述。霍金的这个很可能成为他最杰出作

品的最新模型，其值得惊讶之处倒不在于是出自一位残疾人，而在于是出自一位已经40岁出头的人，这对一个要取得新突破的数学家来说已是很晚的年龄。即使爱因斯坦在完成广义相对论时也只有30多岁，而此后他就再没有取得什么有重大意义的科学成果。

霍金所提出的模型仍在发展之中，还没有得到像他较早的工作那样的承认。但它击中了关于时间、空间和物质起源存在已久的疑难的核心，它也很显然是现有的对创世时刻最完整最有条理的描述。

霍金的宇宙

按照广义相对论，在时间的起始即创世时刻必定有一个奇点。但是像已有的所有物理理论一样，广义相对论对于比普朗克时间即 10^{-43}秒更早的时间是不适用的。虽然稳恒态思想的某个变种，即一个均匀的超宇宙或是某种整体混沌再加上暴胀，也有可能提供一种方式来产生像我们宇宙这样的膨胀着的局部时空区域，但是如果能建立一个数学模型即一组方程式，能对我们的宇宙作出自容的描述，尤其是能避免 $t=0$ 奇点的窘迫，那当然会更令人满意得多。这就是霍金探索我们的起源难题的基本出发点，也是他试图在一个有效的宇宙模型里至少是部分地把广义相对论和量子理论结合起来的基本出发点。[8]

但是，什么形式的量子物理学适宜于描述宇宙整体呢？要记住量子理论对一个粒子或一个系统如何从状态A到达状态B可是什么都不能告诉我们。对量子物理的正统解释是哥本哈根解释。它说的是，当我们不看一个系统时，它是存在于某种所有可能状态的超位置，而对该系统作测量或是看一看的行为，都会使“波函数收缩”到这些可能状态之一，而这种状态只是依据概率来选取的。在我们停止测量或看该系统时，它在量子意义上又从那个确定状态散开成一种所有状态

的新超位置，只是到下次被测量时才再收缩，或许是收缩成不同的状态。这种解释作为一种实用工具来计算原子和亚原子粒子（其实还有分子）的行为非常有效。但它很难为常识所接受，而且在试图把哥本哈根解释运用于整个宇宙时真正遇到了困难。

当然，我们可以设想宇宙能用量子力学波函数来描述，哪怕是永远也不能希望写下描述整个宇宙的“波函数”的方程式。但由于宇宙按其定义是包含我们所能知道的一切，也包含我们自己，那么就没有人在宇宙之“外”来观测它并导致它收缩到一种可能的量子态。计算描述宇宙（或任何包括观测者在内的系统）行为的概率的正确途径，是采用量子物理的另一种解释，即所谓多种世界解释，其中该系统的所有可能波函数的效应原则上能用费恩曼的路径积分（或“历史求和”）方法计算并相加起来，从而得出对该系统以及它如何从状态 A 到状态 B 的整体数学描述。

在大多数情况，由这两种途径得出的对量子物理问题的解答是相同的，几乎没有什么物理学家愿意再去为多种世界解释烦神，因为他们在成长过程中受的是哥本哈根解释的教育，他们已经习惯了一种违反常识的思想，很难再去适应第二种非常识思想。但如霍金曾在好几个场合指出的（包括在 ESO/CERN 会议上的报告），多种世界解释与历史求和的结合是寻求对宇宙的量子描述的唯一途径，在这个意义上“历史求和”确实就是把宇宙所有可能的演化方式加在一起。我们当然连详细计算宇宙的一种“历史”都不可能做到。但我们能为宇宙选择一组起始条件，也就是边界条件，我们（或者说是霍金及其合作者）还能计算简单形式的宇宙的演化，那种宇宙只包含一对场（一个表示引力，一个表示物质）。希望是这个被霍金称为微超空间的简单模型能够与真实宇宙有足够的相似，从而使他能推导出宇宙演化的主要特

征。这个希望看来是能实现的。

霍金选择的宇宙边界条件只有在量子物理与广义相对论结合时才可能出现。广义相对论说对 $t=0$ 必定有一个奇点。而经典物理也说过原子中应该有一种奇点，意思是电子不可能停留在围绕核的轨道上，而是必定会落到核上。量子力学解释了为什么原子能够稳定存在，它也使从宇宙学中移去 $t=0$ 的奇点有了指望。用物理语言讲，我们可以把时间起源设想成是被量子不确定原理抹开在 10^{-43} 秒的时间上，因而并没有一个单一的创世时刻。这样一来，用一个由广义相对论和量子物理的适当数学结合来建立的物理模型，就有可能把四维时空作为一个像球面或者地球表面那样的闭合曲面来描述。

以前讲过由三维空间组成并随着时间流逝而膨胀的闭合曲面。但现在必须也这样来考虑时空，而不只是空间。将比拟再作延伸，空间就不是由一个面而是由一条线来代表，比如说可以选取为在代表时空网的球面上的纬度线。时间就能由从极点开始沿一条经线的“距离”来代表，比方说如果从北极开始朝赤道移动，代表空间的圆圈(纬度线)就随着“时间”的流逝即向赤道趋近而变大。这样一个宇宙模型是完全自容的。时间或空间都没有边缘，也没有奇点。这是可能描述宇宙的最简单的几何，这也只是因为量子效应改变了相对论规则才能存在的几何，相对论规则本来是要求在时间的起始必定有奇点的。霍金强调，他所提议的这种宇宙状态只是一个提议。他给出的宇宙边界条件是“它没有边界”，即没有边缘，没有奇点，时间和空间都无始无终。惊人的事情是，所有可能的边界条件中最简单的这一种却使他对宇宙作出了完全合理的描述，这个描述与我们在周围看到的别无二致。

霍金用量子物理的历史求和方法来检验这个模型。从原则上讲，

就是把适合边界条件的所有可能历史的效应都加起来，也就是把大小有限、没有边界的所有可能宇宙的效应都加起来。但在实际过程中，他就不得不作许多简化假设，把他的模型宇宙压缩到只有前面提到的两种基本场。但当他这样做并进行路径积分时，他发现绝大多数历史都由于相邻路径的干涉而抵消，正如对原子中电子的轨道作相应计算时的情况那样。只有很少几种可能的历史被增强，因而就有着很高的发生概率。它们组成一个高概率历史族，有几个重要的共性。其一是它们在所有三个空间方向上都一致地膨胀；其二是它们中的每一个都膨胀到一定尺度，接着就收缩回到极高密度状态，就像我们的宇宙在普朗克时间的状态那样，然后再次膨胀。每次膨胀和收缩的循环都与前一次完全一样。宇宙在这种连续的循环中不会膨胀得更大或更小，而总是膨胀到一样大。更好的是，霍金模型中两种场的相互作用能造成初始的极快膨胀状态，就是暴胀，然后才是物质开始主宰宇宙并使它转到我们今天在自己宇宙里看到的这种平稳膨胀。

这些允许的宇宙历史中任何一个都可以是对我们宇宙的很好描述，也就是说我们宇宙是一个无边界无奇点的闭合系统，处于膨胀、收缩、再膨胀的永久循环之中。容易理解为什么霍金为他的模型所给出的可能性而兴奋。按照这幅图景，宇宙必定是闭合的，而暴胀又使它接近于平坦，从而解决了以前也由暴胀来解决的那些老问题。但模型还给出了一些奇特的美妙的结果。首先是在允许的历史族里还有其他宇宙，在一定意义上与我们并肩地（在超空间中的隔壁）经历着它们的膨胀和收缩循环。但是我们没有任何办法能知道它们，更不要说同它们联系了。由于量子物理的作用方式，一旦我们作测量或进行实验，我们将得到与一个量子态相一致的结果，这个量子态也就是一个描述“我们的”宇宙及其中包括我们自己在内的一切的波函数。如果

别的智慧生物占有另一个量子态，对应着第二个高度可能的宇宙波函数，他们将作出自己的观测并总是得到相应于那个波函数的结果。除了抵消一些波函数和加强其他一些之外，量子态并不相互作用，也就是没有干涉，实验结果总是与方程的一个或另一个“经典”解一致。尽管霍金模型说超空间里并存着许多世界而我们只能知道其中一个，我们毕竟能对自己这个宇宙的命运和时间的本质作出新的洞察。

再来看霍金的4维平滑球面宇宙模型。从北极开始不断向外增大的纬度环代表宇宙空间的膨胀，而北极本身就代表大爆炸即创世时刻。但是在极点上没有奇点，我们只是从那里开始量度时间。同样道理，地球的北极是我们开始量度纬度的地方(虽然我们通常是把赤道定义为纬度0°，但同样可以从极点开始量度)，这并不表示在极点有奇点。没有比北极更北的地方，但这并不表示在那里有一个空间边界。没有比$t=0$更早的时候，但这并不表示时间从那时开始。创世时刻$t=0$现在只是一个用于量度时间的合适标记。

霍金首次公布这些思想是在1981年梵蒂冈的一次宇宙学会议上。与会的物理学家和数学家受到了教皇的接见，教皇说他们研究宇宙在创生之后的演化是很恰当的，但是时间起始的难题属于宗教而不是科学的范畴，因为那是上帝的工作。[9] 也许教皇的劝告是太得体了，以至于不好向他指出，霍金的宇宙模型已经免去了时间起始的奇点，因而也免去了教皇归功于上帝的工作。或者就是霍金在会议上报告的含意太深刻了，以至于没有被充分领会。

霍金的模型去掉了宇宙“起始”时麻烦的奇点。但若把与球面上纬度线的类比进一步延伸，在赤道上会发生什么呢？如果在球面上画出一系列越来越往南的纬度线，线圈就会增大，也就是宇宙在膨胀，直到赤道为止，此后线圈就会越来越小，最终收缩到南极上而不复存

在。过赤道后纬度圈的变化就相当于霍金宇宙的收缩阶段，这对认识时间的本质有重要意义。时间的存在是由于宇宙从低熵态到高熵态的演化。宇宙从大爆炸的有序态膨胀到遥远将来的无序态。看来可以自然地猜想，宇宙在其生命的收缩阶段将从无序态演化到有序态，熵的流动逆转，把宇宙“钟”的发条上紧得恰到好处，以保证下一次膨胀和收缩的循环也与该特定宇宙史上以前所有的循环精确一样地进行。三维宇宙在小的时候是有序的，大的时候是无序的。为什么它在小的时候就该有序呢？为什么在宇宙收缩时熵会被迫倒流呢？这也许简单地是因为在较小的宇宙里没有无序的“余地”。沿着赤道或是靠近它的纬度线可以扭动而仍然很好地近似于纬度线。而紧靠北极或南极的纬度线就不可能扭动而又不越过极点。在普朗克极限即 $t = 10^{-43}$ 秒时，宇宙是如此之小，因而必定是很均匀的。如果这幅图景正确，这就是为什么由熵定义的时间之箭与宇宙随时间流逝而变大的时间之箭是一样的缘故。那么在收缩的宇宙里这两枝时间之箭会步调不一吗？霍金说未必。他在1983年所作的一个系列讲座里曾说过：“箭是否与由膨胀定义的时间方向一致尚有待查明。”[10]他的意思是可以设想两枝箭一起倒转，因而仍然相互一致。但这会引出一些古怪的后果，曾由天文学家戈尔德(稳恒态模型的提出者之一)在20世纪60年代讨论过。

戈尔德的猜测可没有霍金对宇宙的数学描述那么高的出发点，而只是简单地从宇宙闭合和有限，因而有一天会再收缩这种可能性开始。他试图给出熵流逆转的收缩宇宙的热力学。[11]恒星不再是把热量辐射到空中，而是吸收电磁辐射，来自空中所有方向的电磁辐射聚集到就近的各个恒星上。在恒星内部，进入的辐射会驱动把氦转变成氢的核过程，而在像我们地球这样的行星上，冰块会辐射出热并增大，

生物会从老年“长”得年轻。这似乎是一派胡言，其实不过是把我们所看到的周围世界用时间反演的语言来描述。真实存在的是从一种状态到另一种的变化，至于定义那个变化是“向前”或“向后”的标记则是任意的。戈尔德说，如果熵反向流动，那么使智慧生物成其为智慧生物的思维过程也是如此。任何生活在我们称为宇宙收缩阶段的智慧生物，将会相对于我们的思维方式“反向”思维，因而将仍然“看到”热从高温物体流到低温物体，并将推断出宇宙是在膨胀而且最终会转到收缩状态，而他们认为的收缩正是我们认为的膨胀！对智慧生物来说，不仅霍金模型没有边界，而且宇宙的每个量子历史的每个循环也没有终结，而只有两个相区别的起始。

当然，对“赤道”上即宇宙从膨胀转为收缩时会发生什么还是有问题的。戴维斯推测，这些问题可以在这样一个模型中解决，即时间之箭在整个一次膨胀和收缩的循环中是一样的，但在火球阶段会倒转，反正在火球里所有信息都会丧失，于是在下一次膨胀和收缩的循环里时间之箭和熵的流动就倒过来了。这个想法远在霍金的量子宇宙模型之前就被提出了（见《失控的宇宙》），但它向我提示了这样一种简洁的可能性，即在霍金宇宙的一次膨胀和收缩中，也就是“纬度线”从北极到赤道再到南极的一次增大和缩小过程中，时间箭和熵流不变向。宇宙的下一阶段就是从南极开始经过赤道收缩到北极上，而熵流则反向。但别太认真地看待我这个猜想。我曾问过霍金这是否可行，他说：“不。南极是一个正常时空点，这就要求宇宙在收缩到小体积时是平滑和均匀的。也就是说在最大膨胀点上的无序和不均匀状态必须随着宇宙的收缩而变得有序。因此在收缩阶段热力学时间箭必定反向。”在这个课题上，肯定他是头。

撇开猜想不谈，霍金的宇宙模型还有着重要的哲学意义。尽管细

节有待补充，也还大有猜想的余地，他却告诉我们，原则上有可能建立一个单由已知的科学定律就能完整地描述宇宙的数学模型，即使在创世时刻都不需要乞求任何特殊条件。量子物理学是解开宇宙的最后秘密、解释其起始和终结所需要的钥匙。如果说今天对暴胀的认识状况相当于20世纪40年代之对大爆炸理论，那么把里斯所做的这个比拟延伸，就有理由说霍金的量子宇宙模型现在是相当于20世纪20年代初的量子物理学的地位，也就是海森伯、薛定谔等人建立起完整的自洽理论之前的地位。物理学家正在摸索并有可能在20世纪结束前发现(或发明)的统一理论，应该能与霍金模型结合起来，解释宇宙中已经发生和可能发生的一切现象。这样结合起来的理论至少在原则上将能够预言任何事件发生的可能性，尽管实际计算会过分复杂，以至于方程式只有在最简单情况才解得出来。

我们对大爆炸和大爆炸之前直到创世时刻的探索到此结束了。霍金模型展示了把广义相对论和宇宙学结合成为一个大创世理论的前景，并且指出我们已经知道了所有的基本物理定律。不需要乞灵于奇迹或是新物理学，就能解释宇宙是从何而来。霍金自己曾预言了“理论物理学的终结”，意思是一个也许建立在N=8超引力基础上的令人满意地统一一切的理论，可能已经在望并且在20世纪结束之前能够达到。当然仍有大量工作留待物理学家去做，诸如补充宇宙演化的细节之类。也许更准确地说应该是，霍金已经指出了一个终点，但不是对物理学而是对形而上学。现在已经可以对“我们从何而来？”的问题作出很好的科学回答，而不必乞求上帝或是宇宙在创生时刻的任何特殊边界条件。1981年的梵蒂冈会议使形而上学家失业了。其余的一切都随之变得不重要了，形而上学的道路终结的地方肯定也是本书终结的好地方。

注　释：

1. 英国物理学家戴维斯确曾提出，正是暴胀给出了我们宇宙中时间的意义。最初的混沌是一种高熵态，而暴胀后的宇宙是很低引力熵的状态。低熵意味着热能够流动，因而随着时间流逝宇宙看上去会不一样，于是时间流逝的概念才有了意义。用戴维斯的话说，宇宙是由暴胀“上紧发条”的。“宇宙此后的历史就是由引力聚集（星系→恒星→黑洞）和核合成（氢→氦→铁）来松开发条。这两个演化链合起来就能说明世界上所有观测到的宏观时间不对称，并且在我们周围留下了时间之箭的明确印记。” *Nature*，Vol. 301，p. 398；亦见 Vol. 312，p. 524。

2. 应该说，在无限的总宇宙里任何可能的事情都会在无数地方发生无数次。

3. *Nature*，Vol. 246，p. 396. 苏联的福明（P. I. Fomin）大约同时也独立提出了一个类似的想法，并以“预印本”的形式散发，直到1975年才正式发表。

4. 这个“一无所获”的基本思想至少可以推回到特赖恩的文章之前30年。伽莫夫在自传《我的世界线》里，讲了二战期间他在华盛顿的美国海军部军械局当顾问的情况。伽莫夫没被允许参与研制原子弹，因为他生于俄国，而且老喜欢对朋友们说自己20岁时在红军里当过上校。尽管知道他爱讲大话和开玩笑，曼哈顿计划里那些负责安全工作的人可不敢马虎，于是他战时就留在华盛顿。

但伽莫夫有一项任务，每两周一次把一包文件送给在普林斯顿的爱因斯坦。这些文件虽然形式上是保密的，其实与核武器无关。其中所讲的是各种新武器设想，海军部希望爱因斯坦作出评论。据伽莫夫说，不管那些设想是多么稀奇古怪，爱因斯坦几乎总是予以好评。有一天他俩一起从爱因斯坦家步行到高等研究院，途中伽莫夫提起乔丹（Pascual Jordan）有一个新想法。乔丹是量子物理学的创立者之一，他同海森伯和玻恩（Marx Born）一起于1925年建立了第一种形式的量子力学，即所谓矩阵力学。但他的新想法在20世纪40年代时看来可不是属于这一类的，那只是物理学家们喜欢在喝咖啡时或是走在普林斯顿的路上时谈谈的古怪想法之一。伽莫夫边走边告诉爱因斯坦的是，乔丹认为恒星可以从虚无中产生出来，因为在零点上它的负引力能与正的静止质能数值上是相等的。

伽莫夫写道：“爱因斯坦站住不走了，那时我们正在过马路，好几辆汽车不得不刹住，以免撞倒我们。”

正是那个使爱因斯坦停步的思想在20世纪80年代受到了像特赖恩这样的研究者的高度重视，并且被运用到了整个宇宙，而不只是单个恒星。

5. 维连金不同寻常的经历值得作一简短介绍。他于1949年出生在苏联的哈尔科夫，1971年从国立哈尔科夫大学毕业。但他告诉我，因为他是犹太人，此后就得不到研究职位，只好先在军队服务，后来靠各种临时工作为生（他说自己最喜欢干的是在动物园值夜）。1976年他移民到美国。但在那5年里他一直用空余时间学物理，而且学得那么好，1977年即到美国后一年就从纽约州立大学布法罗分校获得了博士学位。

6. 不只是他关于相对论的工作。我第一次去剑桥是在1967年，是去听关于瓦戈纳、福勒和霍伊尔小组原初核合成研究成果的报告，这在当时是很新的思想。在一间挤满人的教室里有一些英国最好的物理学家和天文学家，在那里像我这样的研究生该做的就是保持沉默和记笔记，但是听众中最尖锐的问题是一个我以前从没见过的年轻人提的，他显得稍有点口吃，但显然对所讨论课题的了解已是一流水平。当然，他就是霍金。

7. 这个奖被物理学家公认是最高荣誉，高于诺贝尔奖。

8. 应该承认几乎没有什么物理学家对霍金的探索表示完全赞同。他不得不作许多简化假设，而别人并不总是赞成他处理方程式的方式。但是作为模型基础的物理原理是非常清楚和直截了当的，正是这使我相信霍金是在正确的道路上。方程式的细节可能会改变，但我不相信基础的物理原理会改变。

9. 约翰·保罗二世(John Paul Ⅱ)在1981年会议上的原话是："任何关于世界起源的科学假说，诸如整个物理世界都从一个原初原子派生出来这样的假说，都留下了有关宇宙起源的悬而未决的问题。科学自身不能解答这样的问题，这里需要的是高于物理学和天体物理学的被称为形而上学的知识，而首先需要的是来自上帝启示的知识。"他接着引用了前任教皇庇护十二世(Pius XII)在1951年讲的有关宇宙起源问题的话："等待来自自然科学的答复是徒劳的，恰恰相反，科学会承认所面临的是一个不可解开的谜。"不到40年"徒劳等待"就结束了。许多宇宙学家不再认为谜是不可解开的，而是觉得科学能够解决关于宇宙起源的形而上学难题。有意思的是，霍金的宇宙正是在那个保罗二世强调上帝伟大作用的会议上亮相的，它清楚地指出了对这个最大的形而上学难题作出最终科学解答的途径。

10. 在法国乌什暑期讲习班上，这里对霍金宇宙的介绍大部分取材于该讲座已发表的讲义，加上他在ESO/CERN会议上的报告，还有他在1984年给我的几篇论文的复印件。

11. 对戈尔德模型一个很好很简短而又易于理解的评述见于 Paul Davies, *The Runaway Universe*(London: Dent,1978),p.194。据我所知，戈尔德最早提到这项研究是1964年在康奈尔的题为"时间的本质"的会议上。但戴维斯告诉我他现在已追踪到一篇1958年的论文。

参 考 文 献

Abell, George. *Exploration of the Universe*, fourth edition. Philadelphia: Saunders, 1982.

Bernstein, Jeremy. *Three Degrees Above Zero: The Bell Labs in the Information Age*. New York: Charles Scribner's Sons, 1984.

Bonnor, William. *The Mystery of the Expanding Universe*. New York: Macmillan, 1964.

Chandrasekhar, Subrahmanyan. *Eddington*. Cambridge, Eng.: Cambridge University Press, 1983.

Davies, Paul. *The Forces of Nature*. Cambridge, Eng.: Cambridge University Press, 1979.

Davies, Paul. *Superforce*. London: Heinemann, 1984.

Ferris, Timothy. *The Red Limit*, revised and updated edition. New York: Quill, 1983.

French, A. P. (editor). *Einstein: A Centenary Volume*. Cambridge, Mass.: Harvard University Press, 1979.

Fritzsch, Harald. *Quarks*. London: Pelican, 1984.

Gamow, George. *The Creation of the Universe*. New York: Viking Press, 1952.

Gamow, George. *Mr. Tompkins in Paperback*. Cambridge, Eng.: Cambridge University Press, 1967.

Gamow, George. *My World Line*. New York: Viking Press, 1970.

Gibbons, G. W.; S. W. Hawking; and S. T. C. Siklos (editors). *The Very Early Universe*. Cambridge, Eng.: Cambridge University Press, 1983.

Gribbin, John (editor). *Cosmology Today*. London: New Science Publications/IPC Magazines, 1982.

Gribbin, John. *In Search of Schrodinger's Cat*. New York: Bantam; London: Corgi, 1984.

Harrison, Edward. *Cosmology*. Cambridge, Eng.: Cambridge University Press, 1981.

Hoyle, Fred. *Galaxies, Nuclei and Quasars*. London: Heinemann, 1965.

Hubble, Edwin. *The Realm of the Nebulae*. New York: Dover, 1958.

Judson, Horace Freeland. *The Eighth Day of Creation*. London: Cape, 1979.

Layzer, David. *Constructing the Universe*. New York: Scientific American Books, 1984.

Narlikar, Jayant. *The Structure of the Universe*. London: Oxford University Press, 1977.

Novikov, I. D. *Evolution of the Universe*. Cambridge, Eng.: Cambridge

University Press, 1983.

Pais, Abraham. *Subtle Is the Lord*. London: Oxford University Press, 1982.

Pears. D. F. (editor). *The Nature of Metaphysics*. London: Macmillan, 1957.

Pickering, Andrew. *Construction Quarks*. Edinburgh: Edinburgh University Press, 1984.

Polkinghorne, J. C. *The Quantum World*. London: Longman, 1984.

Sciama, Dennis. *Modern Cosmology*, revised edition. Cambridge, Eng.: Cambridge University Press, 1973(reprinted in paperback, 1982).

Silk, Joseph. *The Big Bang*. San Francisco: W. H. Freeman, 1980.

Sutton, Christine. *The Particle Connection*. London: Hutchinson, 1984.

Weinberg, Steven, *The First Three Minutes*. London: Deutsch, 1977.

Wright, Thomas. *An Original Theory of the Universe*. London: Macdonald, 1971.